DOCUMENT RESUME

ED 464 365 CS 511 125

AUTHOR	O'Neal, Marcia R.; Spor, Mary W.
TITLE	Evaluation of the Alabama Reading Initiative, 1999-2000.
INSTITUTION	Alabama Univ., Birmingham. School of Education.
SPONS AGENCY	Alabama State Commission on Higher Education, Montgomery.
PUB DATE	2002-00-00
NOTE	84p.; Prepared with contributions from Scott Snyder. Prepared by Center for Educational Accountability, University of Alabama at Birmingham. Evaluation Project funded by Dwight D. Eisenhower Professional Development Program. For executive summary, see CS 511 133. Appendixes are not available from ERIC or on the Web site. For "Evaluation of Year 2," see ED 460 144.
AVAILABLE FROM	Alabama Reading Initiative, Alabama Department of Education, P.O. Box 302101, Montgomery, AL 36130-2101. Tel: 334-353-1389; Web site: http://www.alsde.edu.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Administrator Role; Elementary Secondary Education; Evaluation Methods; Higher Education; Literacy; *Preservice Teacher Education; Professional Development; Program Effectiveness; Reading Achievement; *Reading Improvement; Reading Instruction; *Reading Programs
IDENTIFIERS	Alabama Reading Initiative

ABSTRACT

A study examined the effectiveness and the impact of preservice instruction of the second year of the Alabama Reading Initiative (ARI), which is designed to achieve 100% literacy by targeting reading performance of beginning reading and first-grade students, to expand the reading power for second- through twelfth-grade students, and to intervene for struggling readers at all grade levels. In its second year of implementation, approximately 27,700 students, 2,354 teachers, 81 principals, 75 reading specialists, 64 higher education partners, and 221 preservice teacher education programs were directly involved in the ARI. The evaluation used information from student achievement data from surveys completed by teachers, reading specialists, principals, higher education partners, and higher education reading faculty. Both qualitative and quantitative data were analyzed. A pretest-posttest design was used to examine change over time for student outcomes. Results and recommendations regarding delivery of the ARI are: a large percentage of principals, teachers, reading specialists, and higher education partners reported the ARI had substantial positive impacts on student literacy; the principal should receive sufficient support, direction, and encouragement because their leadership role was related to gains in reading scores; the visibility and involvement of the reading specialist and higher education partner influenced teachers' attitudes toward the helpfulness of these individuals; higher education faculty in the program are often confronted with inadequate or competing reward structures; teachers involved in the first 2 years of the program felt that continued professional development would help sustain their enthusiasm; the Stanford Achievement Test, Level 9 is not the most appropriate measure for monitoring progress toward program goals; tests currently used to assess early literacy skills of Kindergarten through second-grade students are inappropriate for use as tools

Reproductions supplied by EDRS are the best that can be made
from the original document.

for evaluating gains associated with ARI; and case studies of particularly effective schools should be conducted. (Contains 25 tables and a figure of data.) (RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 365

EVALUATION OF THE ALABAMA READING INITIATIVE
1999-2000

Prepared for the
Alabama Commission on Higher Education

by the

Center for Educational Accountability
University of Alabama at Birmingham
233 Education Building
901 13th Street South
Birmingham, Alabama 35294-1250
Telephone: (205) 934-7598
Fax: (205) 975-5389

Project Directors:
Marcia R. O'Neal, Ph.D.
Mary W. Spor, Ph.D.

Other Contributor:
Scott Snyder, Ph.D.

This evaluation project was funded in part by a federal grant under the Dwight D. Eisenhower Professional Development Program administered by the Alabama Commission on Higher Education.

Opinions and findings expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education or the Alabama Commission on Higher Education, and no official endorsement by either of these agencies should be inferred.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CS 511 125

PART 1: INTRODUCTION

Goals of the Alabama Reading Initiative

In 1996 the Alabama State Board of Education appointed a 25-member group to review research on reading and to develop a plan to target K-12 literacy in the state. This group, called the Alabama Reading Panel, was composed of representatives from classrooms, colleges and universities, business and industry, and grassroots support groups. The plan that the Panel developed, known as the Alabama Reading Initiative (ARI), is a statewide effort to improve reading instruction significantly and for public school students ultimately to achieve 100% literacy. ARI is designed to target reading performance of Alabama students on three fronts: 1) beginning reading for kindergarten and first-grade students, 2) expansion of reading power for students in Grades 2 through 12, and 3) intervention for struggling readers at all grade levels.

The Alabama Reading Initiative Process

The ARI seeks to accomplish its goals through the implementation of research-based reading instruction at Literacy Demonstration Sites (LDSs). These sites were selected by State Department of Education (SDE) staff members and designees who traveled the state, visiting schools that had applied to participate in the program. All public schools in the state are eligible to become LDSs. Selection is based on the candidate school's knowledge of ARI's goals and program content and its commitment to professional development in pursuit of 100% literacy. Establishment of the LDSs allows for in-depth professional development at the school level by involving the principal and at least 85% of the faculty from each site in two weeks of training during the summer. The Summer Institutes are presented regionally by ARI-trained reading professionals using learning modules developed by ARI.

Unique elements of the ARI model include 1) its three key goals 2) its all-inclusive K-12 focus, 3) its requirement that a minimum of 85% of the faculty attend the

training, 4) its requirement that the principal participate in the same training as the faculty and lead the faculty during 10 hours of faculty meetings during the Summer Institute, and 5) its formation of partnerships between the LDSs and professional educators from Alabama's Institutions of Higher Education (IHEs).

The Alabama Reading Initiative and the Higher Education Partnership

The ARI is a collaborative model in which IHE faculty have been partners from the outset. These Higher Education professionals have served on the Alabama Reading Panel, helped to develop instructional modules for the Summer Institutes, trained both the presenters and their trainers for the Summer Institutes, and presented learning modules at the Institutes. They also participated in the Advanced Study Group, which during Year 1 was dedicated to reviewing current research and practice in-depth. An important by-product of the active involvement of partner IHE faculty in ARI is the opportunity for alignment of pre-service reading education with the research-based, professional development provided at the Summer Institutes.

Implementation of the Alabama Reading Initiative

Implementation of ARI first took place at 16 schools named as Literacy Demonstration Sites for the 1998-1999 school year. For the purpose of this study, these sites are grouped as Cohort A. Sixty-five additional schools became LDSs for the 1999-2000 school year, grouped as Cohort B.

Grassroots support for the ARI model has attracted substantial private funding for the Summer Institutes as well as for other key elements of the program. During ARI's first year, 1998-1999, no financial resources were available to support involvement of higher education partners; therefore, their participation was entirely on a voluntary basis. During the second year of ARI, 1999-2000, legislative funding was provided for partner IHE faculty to visit the 81 LDSs on a monthly basis. State funding also was provided for 75 full-time reading specialists (trained through ARI) to serve the 81 LDSs. Each

reading specialist supported ARI implementation by working on a daily basis at the LDS to help its struggling readers and to coach its faculty relative to implementation of Summer Institute learning.

In its second year of implementation, approximately 27,700 students, 2,354 teachers, 81 principals, 75 reading specialists, 64 higher education partners, and 21 pre-service teacher education programs in Alabama's IHE were directly involved in the ARI.

Of the 81 LDSs participating in the second year of the ARI (Cohort A and Cohort B together), 5 were primary schools; 56 were elementary schools; nine were middle schools; four were high schools; three were K-12 schools; three were K-8 schools; and one a K-7 school. Twenty-four schools were from small school systems (3,900 students or less), and the other 57 schools were from systems classified by the Alabama State Department of Education as large (more than 3,900 students). School populations ranged from 188 students to 1,153 students. Socioeconomic status, as measured by the percentage of students receiving free lunch, ranged from 2.56 to 99.10 percent. In 30 of the 81 schools, more than half of the students received free lunch.

PART 2: THE EVALUATION PROJECT

Purpose

As with the Year 1 evaluation[1], the evaluation of Year 2 was crafted to examine the Alabama Reading Initiative from two perspectives. In the K-12 Component, ARI is evaluated for the purpose of improving its effectiveness. In the Pre-Service Component, ARI is evaluated for the purpose of determining its impact on pre-service instruction.

The K-12 Component

A primary purpose of the Alabama Reading Initiative (ARI) Year 2 evaluation is to gain an understanding of the factors that could be linked to increased student achievement in reading. With this understanding, ARI can adjust its requirements and procedures to reach its goals of 100% literacy and of expanding to all schools in the state. In order to achieve successful statewide expansion, the ARI needs to know which activities, requirements, and/or features are essential and which are optional; which are strong and which are weak; which should be replicated and which should be adjusted.

Evaluation of The K-12 Component is therefore embodied in four vital questions:

(1) To what extent are ARI schools making progress toward 100% literacy?

(2) Which ARI schools are making progress toward 100% literacy and which are not?

(3) What factors are related to school outcomes?

(4) Why are some ARI schools making more progress than others?

The Pre-Service Component

The long-range solution to Alabama's pursuit of 100% literacy lies, at least in part, with teacher preparation programs. Teacher preparedness to instruct all public school students to learn to read, to expand their reading power, and to provide

[1] See "Evaluation of the Alabama Reading Initiative 1998-1999," prepared for the Alabama Commission on Higher Education by the Center for Educational Accountability, University of Alabama in Birmingham, Birmingham, AL.

intervention for struggling readers depends in part on the preparation of teachers and administrators in pre-service programs.

Two questions are posed in evaluating the pre-service component:

(1) To what extent are the elements of ARI reflected in pre-service teacher education programs throughout Alabama?

(2) What ARI factors are related to change in pre-service teacher education programs?

Management

This evaluation was approached from an internal/external perspective. The external evaluation Project Director, Dr. O'Neal, from the University of Alabama at Birmingham Center for Educational Accountability (CEA), was responsible for data analysis and interpretation to ensure the integrity of the findings. The internal evaluation Project Director, Dr. Spor, from the University of Alabama at Huntsville, and representatives from the ARI staff participated in the design, development, and execution of the evaluation.

Structure

Evaluation of Year 2 of the ARI is a collaborative effort conducted by representatives from the University of Alabama at Birmingham (UAB) and the State Department of Education. The evaluation consists of two major components. The K-12 component examines student, teacher, reading specialist, principal, and higher education partner outcomes at the 81 Literacy Demonstration Sites (LDSs) and the factors that may influence student achievement. The pre-service component examines changes in pre-service teacher education practices at the 32 public and private Institutions of Higher Education (IHEs) in Alabama that house teacher education programs. Of these, 20 were mentor IHEs selected by LDS schools. A list of the LDSs and IHEs can be found in Appendix A.

The ARI evaluation has been funded by a federal grant under the U.S. Department of Education Dwight D. Eisenhower Professional Development Program (P.L. 103-382) administered by the Alabama Commission on Higher Education (ACHE). The evaluation has been further guided by input from an Evaluation Oversight Committee composed of representatives from private industry, ACHE, the Alabama State Department of Education (SDE), institutions of higher education, the A+ Foundation, and LDSs (central office administrators and building principals). Evaluation Oversight Committee members are listed in Appendix A.

The evaluation uses information from student achievement data and from surveys completed by teachers, reading specialists, principals, higher education partners, and higher education reading faculty. Both qualitative and quantitative data were analyzed. A pretest-posttest design was used to examine change over time for student outcomes.

Instruments

Both achievement data and survey responses were used for the K-12 component of the evaluation, and survey responses were used for the pre-service component. Achievement data used in the evaluation include results from the *Stanford Achievement Test, Ninth Edition* for students in Grades 3-11. Survey instruments were developed collaboratively by Drs. O'Neal and Spor with feedback from the Oversight Committee. Both rating-type and open-ended items were included in instruments distributed to teachers, principals, reading specialists, higher education partners, and higher education reading faculty. Responses from teachers, principals, reading specialists, and higher education partners comprised survey data for the K-12 component. Responses from higher education partners and higher education reading faculty comprised data for the pre-service component.

Data Collection and Processing

Data collection, analysis, and presentation also were a collaborative effort. The Alabama State Department of Education provided student achievement data.

Data collection relative to the K-12 survey instruments was facilitated through regional meetings with ARI principals and/or reading specialists where the evaluation purpose and design were presented and survey instruments were provided and discussed. Principals, reading specialists, and teachers completed surveys independently and mailed them back to CEA in postage-paid envelopes. Principals appointed a teacher whom faculty trusted to be in charge of a faculty meeting where surveys were distributed and completed. Each teacher placed his/her survey in a sealed envelope and gave it to the teacher in charge, who placed each sealed envelope into a large postage-paid envelope and mailed it to the CEA.

Higher education partner surveys, which solicited data for both K-12 and pre-service components, were mailed directly to higher education partners. Higher education reading faculty surveys, soliciting data for the pre-service component, were mailed to deans or department chairs at 32 public and private institutions of higher education in Alabama, who were asked to distribute the surveys to their reading faculty. Completed surveys were mailed to the CEA.

CEA faculty and staff analyzed student achievement data, prepared sufficient copies of all survey instruments, distributed surveys to ARI schools, received completed surveys, entered and analyzed survey data, and prepared tables and narratives of results for all data. The process of preparing tables and narratives included ongoing feedback from the SDE ARI staff and consultants as well as feedback at designated times from the Oversight Committee.

PART 3: THE K-12 COMPONENT

Key Issues

The K-12 component of the evaluation addresses student, teacher, reading specialist, principal, higher education partner, and site outcomes at the 81 LDSs. The focus of the K-12 component of the evaluation is on the following questions:

(1) To what extent are ARI schools making progress toward 100% literacy?

(2) Which ARI schools are making progress toward 100% literacy and which are not?

(3) What factors are related to school outcomes?

(4) Why are some ARI schools making more progress than others?

Evaluation Methods - Outcomes

In order to answer these questions, the following were examined:

- data from the *Stanford Achievement Test, Ninth Edition (Stanford 9)* for Grades 3 through 11, and
- information from survey instruments containing both quantitative and qualitative questions.

The qualitative questions in the surveys concerned teachers' perceptions of the impact of the ARI upon their practices and student learning. The survey instruments also were used to probe and quantify certain collateral outcomes (discipline referrals, special education referrals, and library circulation) that had been reported only anecdotally over the course of the ARI program. Achievement instruments are more fully described in Appendix B. Copies of the survey instruments are included in Appendix C.

Achievement

Originally the evaluation plan called for examination of data from Alabama's statewide kindergarten, first, and second grade assessments. The plan was revised to exclude these analyses for the following reasons:

(1) The assessments are given in the fall of each year, before substantial instruction is implemented in the schools. Therefore, the use of the kindergarten instrument was deemed invalid for assessing the impact of ARI on student achievement in kindergarten.

(2) The assessments are not linked; consequently, changes in a group of students over time could not be examined.

(3) Ceiling effects for some subtests of the instruments diminish the usefulness of results for monitoring differences.

For the *Stanford 9*, evaluation focused on the change in population of stanines 1 through 3 as a reflection of the progress made by struggling readers, the change in population of stanines 5 and above as an indication of the proportion of students reading at or above grade level, and the mean normal curve equivalent (NCE) gains using effect size as a standardized index of change. The results are reported for the following groups:

- Cross-sectional results by school for Cohort A schools for 1998 and 2000 and for Cohort B schools for 1999 and 2000.
- Longitudinal results comparing Grades 3-5 for Cohort A schools in 1998 and 2000 and for Cohort B schools in 1999 and 2000.
- Longitudinal results comparing Grades 6-8 for Cohort A schools in 1998 and 2000 and for Cohort B schools in 1999 and 2000.

- Longitudinal results comparing Grades 9-11 for Cohort A schools in 1998 and 2000 and for Cohort B schools in 1999 and 2000.
- Overall results comparing Non-ARI, Cohort A, and Cohort B. For Cohort A, the comparison of primary interest was the change from 1998 (before ARI training) to 2000. For Cohort B, the comparison of interest was the change from 1999 to 2000.

Collateral Outcomes

In addition to achievement, other school-related outcomes were explored as part of the evaluation. Although schools were not required to maintain or report these data, the following information was solicited from principals:

- number of discipline referrals for the 1997-98, 1998-99, and 1999-2000 school years;
- number of special education referrals for the 1997-98, 1998-99, and 1999-2000 school years; and
- library circulation for the 1997-98, 1998-99, and 1999-2000 school years.

Evaluation Methods – Relationships

The survey instruments were designed by the evaluators in conjunction with the ARI staff to identify possible links between the outcomes cited above and the key features of the Alabama Reading Initiative. Direct relationships may exist, for example, between student achievement and implementation of ARI. Therefore, the following relationships were explored:

- relationships among the student achievement and collateral outcomes,
- relationships between the outcomes and the following ARI process factors:
    - teacher participation in the Summer Institute,
    - principal participation in the Summer Institute,
    - higher education partner involvement throughout the academic year,
    - reading specialist involvement throughout the academic year, and

- implementation of ARI in the classroom.

- relationships among the ARI process factors.

Measurement Instruments

Several methods were used to assess progress toward achieving 100% literacy. In Grades 3 through 11, the primary assessment procedures used to document literacy level were the reading subtests of the *Stanford 9*. The *Stanford 9* is the primary tool used for evaluating school and school system accountability in Alabama.

Four distinct approaches were used for documenting changes in literacy based on the *Stanford 9*. First, progress toward 100% literacy is made when the percentage of struggling readers (defined as students in the lowest three stanines) within a school is reduced. A second metric is found by defining performance at or above the fifth stanine in reading subtests as evidence of reading at or above grade level. When percentages of students within a school scoring at or above the fifth stanine are increased, there is evidence of progress toward 100% literacy. Third, when schools make positive and meaningful increases in averaged normal curve equivalents (NCEs) on the *Stanford 9* across years, there is evidence of progress toward increased literacy. Normal curve equivalents are scores that allow comparisons between different subtests within the *Stanford 9*. Unlike percentile scores, normal curve equivalents can be averaged. Fourth, a different perspective of improvement in literacy is gained by assessing average changes in reading subtest NCEs for groups of students within a school across two or more grades. This type of analysis is described as a longitudinal comparison.

Another method for evaluating progress toward 100% literacy was needed for students in kindergarten through second grade because the *Stanford 9* is only required in Grades 3 through 11. The original intent of the evaluation was to use results from the statewide assessments of students in kindergarten, first, and second grade. However,

for reasons stated earlier, the decision was made to exclude these results from the evaluation.

Stanford 9 results were the primary source of information used to respond to the first two evaluation questions. *Stanford 9* data were available for 15 of the 16 Cohort A schools and 61 of the 65 Cohort B schools. (Five schools were not required to administer the *Stanford 9* because they did not go beyond Grade 2). Pretest (before ARI implementation) and posttest (following ARI implementation) results for Cohort A (1998 to 2000) and Cohort B (1999 to 2000) were examined in several ways. Each set of results is presented and discussed below.

Cross-sectional comparisons of academic achievement within schools (e.g., all students in Grades 3-5 in 1999 in a school compared to all students in Grades 3-5 in the same school in 2000) provide general descriptive information on changes in student achievement.

QUESTION 1:	TO WHAT EXTENT ARE ARI SCHOOLS MAKING PROGRESS TOWARD 100% LITERACY?

Summary of Findings for Question 1

Findings: Overall

- Improvements in average NCE scores in Reading Comprehension, Reading Vocabulary, and Total Reading between 1998 and 2000 averaged 1.05 for Cohort A schools and .24 for non-ARI schools. Improvements in NCE scores across all these reading subtests between 1999 and 2000 averaged .28 for Cohort B schools and .20 for non-ARI schools. These improvements translated into small but positive differences in effect sizes favoring ARI schools for each reading subtest for Cohort A and B.

Findings: Struggling Readers

- Cohort A and Cohort B ARI schools decreased the percentage of struggling readers (students scoring in stanines 1-3) by as much as 10% on the reading components of the *Stanford 9*. Parallel improvements at non-ARI schools over the same period averaged less than half that of the ARI schools.

Findings: Readers At or Above Grade Level

- Cohort A and Cohort B ARI schools increased the percentage of readers at or above grade level (students scoring in stanines 5 and above) by as much as 4.56% on the reading components of the *Stanford 9*. Parallel improvements at non-ARI schools over the same period were consistently smaller than those of the ARI schools.

Detailed Findings for Question 1

Findings: Overall

NCE gains (or losses) were examined for each school. For the purposes of this analysis, all students in a school were included. NCEs were computed for each year,

and the change and effect size associated with that change were calculated. Results for all schools and subtests can be found in tables in Appendix F.

Effect sizes were used to compare differences in school performance across years on all subtests. They permit standardized comparisons of change or difference and are standardized summaries of differences between groups. Effect sizes are determined by computing the differences between means for two groups and then dividing the difference by the average amount of dispersion in the scores (standard deviation). They are regularly used as an index of the practical significance of the differences in scores. Effect sizes can take positive or negative values. When absolute values are between .20 and .40, effect sizes are considered to be small but practically meaningful. Effect sizes between .40 and .70 are considered to be moderate and significant. Effect sizes of greater than .70 are considered to be large.

Cohort A schools that participated in ARI demonstrated greater mean NCE gains between 1998 and 2000 on reading subtests of the *Stanford 9* than did non-ARI schools. Effect sizes for Cohort A schools ranged from .02 to .09, whereas effect sizes for non-ARI schools ranged from .00 to .03. Differences and effect sizes for reading subtests can be found in the table that follows.

NCE Means and Effect Sizes: Cohort A and Non-ARI Schools

	Cohort A Schools				Non-ARI Schools			
Subtest	1998 Mean NCE	2000 Mean NCE	Difference	Effect Size	1998 Mean NCE	2000 Mean NCE	Difference	Effect Size
Total Reading	51.71	52.67	0.96	0.05	49.69	49.87	0.18	0.01
Reading Vocabulary	51.37	53.10	1.73	0.09	49.94	50.48	0.54	0.03
Reading Comprehension	52.02	52.47	0.45	0.02	49.74	49.73	-0.01	-0.00

Cohort B schools that participated in ARI demonstrated little or no difference in mean NCE gains between 1999 and 2000 on reading subtests of the *Stanford 9* than did non-ARI schools. Effect sizes for Cohort B schools ranged from .01 to .02, and effect

sizes for non-ARI schools were .01 for all three reading subtests. Differences and effect sizes for reading subtests can be found in the table that follows.

NCE Means and Effect Sizes: Cohort B and Non-ARI Schools

	Cohort B Schools				Non-ARI Schools			
Subtest	1999 Mean NCE	2000 Mean NCE	Difference	Effect Size	1999 Mean NCE	2000 Mean NCE	Difference	Effect Size
Total Reading	52.70	52.97	0.27	0.01	49.68	49.87	0.19	0.01
Reading Vocabulary	52.68	53.04	0.36	0.02	50.22	50.48	0.26	0.01
Reading Comprehension	52.48	52.68	0.20	0.01	49.59	49.73	0.14	0.01

Findings: Struggling Readers

For the purposes of this evaluation, struggling readers are defined as those students scoring in the bottom three stanines in Reading Comprehension, Reading Vocabulary, and Total Reading. These percentages were computed for each year, and the percentage decrease (or increase) was determined. Results for all schools and subtests can be found in tables in Appendix F.

Cohort A schools that participated in ARI demonstrated greater decreases between 1998 and 2000 in the percentage of students falling in the first three stanines for the reading subtests of the *Stanford 9*. For Total Reading, the ARI schools decreased the population of students in the first three stanines by 9.46%, versus a 1.81% decrease in non-ARI schools. This finding is particularly noteworthy given the evidence that fewer than the statistically expected 23% of the students in 1998 fell within the lowest three stanines. The gains therefore cannot be attributed to statistical artifact. Data for the percentage of change in the Reading Vocabulary and Reading Comprehension subtests are shown in the table that follows.

Percentage Change: Students Scoring in Stanines 1-3
Cohort A and Non-ARI Schools

	Cohort A Schools				Non-ARI Schools			
Subtest	1998 Percentage	2000 Percentage	Difference	Percentage Change	1998 Percentage	2000 Percentage	Difference	Percentage Change
Total Reading	21.81	19.75	-2.06	-9.46	22.25	21.85	-0.40	-1.81
Reading Vocabulary	18.58	16.69	-1.89	-10.19	19.50	18.59	-0.91	-4.68
Reading Comprehension	21.25	19.45	-1.79	-8.45	22.07	22.09	0.02	.09

Cohort B schools that participated in ARI demonstrated greater decreases between 1999 and 2000 in the percentage of students falling in the first three stanines for the reading subtests of the *Stanford 9*. For Total Reading, the ARI schools decreased the population of students in the first three stanines by 7.83%, versus a 2.18% decrease in non-ARI schools. This finding is particularly noteworthy given the evidence that fewer than the statistically expected 23% of the students in 1998 fell within the lowest three stanines. The gains, therefore, cannot be attributed to statistical artifact. Data for the percentage of change in the Reading Vocabulary and Reading Comprehension subtests are shown in the table that follows.

Percentage Change: Students Scoring in Stanines 1-3
Cohort B and Non-ARI Schools

	Cohort B Schools				Non-ARI Schools			
Subtest	1999 Percentage	2000 Percentage	Difference	Percentage Change	1999 Percentage	2000 Percentage	Difference	Percentage Change
Total Reading	18.76	17.29	-1.47	-7.83	22.33	21.85	-.49	-2.18
Reading Vocabulary	15.82	14.81	-1.01	-6.39	19.16	18.59	-.58	-3.01
Reading Comprehension	19.18	17.92	-1.26	-6.56	22.41	22.09	-.32	-1.42

Findings: Readers At or Above Grade Level

Another indication of the success of the ARI is the degree to which the percentage of students falling in the top five stanines for the reading subtests of the *Stanford 9* increases as a result of participation in the ARI model. Results for all schools and subtests can be found in tables in Appendix F.

Cohort A schools showed increases in the percentage of students scoring in the top five stanines for the reading subtests of the *Stanford 9* from 1998-2000. Substantially smaller gains were evident for the comparison schools that did not participate in ARI during the same time frame. For Total Reading, the ARI schools increased the population of students in the top five stanines by 2.18%, versus a 1.28% increase in non-ARI schools. Data for the percentage of change in the Reading Vocabulary and Reading Comprehension subtests are shown in the table that follows.

Percentage Change: Students Scoring at or above the 5th Stanine
Cohort A and Non-ARI Schools

	Cohort A Schools				Non-ARI Schools			
Subtest	1998 Percentage	2000 Percentage	Difference	Percentage Change	1998 Percentage	2000 Percentage	Difference	Percentage Change
Total Reading	62.11	63.47	1.35	2.18	58.94	59.69	0.75	1.28
Reading Vocabulary	63.88	66.79	2.91	4.56	61.05	62.45	1.41	2.31
Reading Comprehension	61.04	62.66	1.62	2.66	58.24	58.65	0.42	.72

Cohort B schools that participated in ARI demonstrated greater increases between 1999 and 2000 in the percentage of students scoring in the top five stanines for the reading subtests of the *Stanford 9*. For Total Reading, the ARI schools increased the population of students in the top five stanines by 1.53%, versus a 0.90% increase in non-ARI schools. Data for the percentage of change in the Reading Vocabulary and Reading Comprehension subtests are shown in the table that follows.

Percentage Change: Students Scoring at or above the 5th Stanine
Cohort B and Non-ARI Schools

	Cohort B Schools				Non-ARI Schools			
Subtest	1999 Percentage	2000 Percentage	Difference	Percentage Change	1999 Percentage	2000 Percentage	Difference	Percentage Change
Total Reading	64.26	65.25	0.98	1.53	59.16	59.69	0.53	.90
Reading Vocabulary	66.50	67.36	0.85	1.28	61.83	62.45	0.63	1.01
Reading Comprehension	63.15	64.31	1.16	1.84	58.15	58.65	0.51	.87

Findings: Longitudinal Results

The cross-sectional reporting methods in the results presented above are the norm for describing academic achievement of Alabama Schools. These cross-sectional

comparisons involve different groups of students for each year. For example, in the illustration above, students who were in Grade 5 in 1999 were not included in the comparison in 2000 because they were in Grade 6. Similarly, a new group of third graders were tested in the year 2000 for the first time.

Researchers often are interested in judging the effectiveness of a program by monitoring the performance of the same group of students across multiple years (or multiple grade levels). This type of analysis of change across time is known as longitudinal research.

The longitudinal comparisons conducted for the study analyze changes in grade level performance (3-4-5, 6-7-8, and 9-10-11 for Cohort A and 3-4, 4-5, 6-7, 7-8, 9-10, and 10-11 for Cohort B). The longitudinal comparisons in this study do not use matched cases. Therefore, the analyses do not necessarily track the same students across time. In addition, five of the fourteen elementary schools in Cohort A could not be included because they lacked Grade 4 and/or Grade 5.

Comparisons of average NCE scores for students in ARI and non-ARI schools in Cohort A (students in Grade 3 in 1998 and Grade 5 in 2000) yielded evidence that greater average gains were found in non-ARI schools in Reading Comprehension and Total Reading. ARI schools demonstrated greater gains in Reading Vocabulary. The longitudinal comparison for middle school (between Grade 6 in 1998 and Grade 8 in 2000) yielded greater gains in ARI schools than in non-ARI schools for every reading subtest. Average NCE scores for the secondary schools (Grade 9 in 1998 and Grade 11 in 2000) yielded greater gains in ARI schools in Total Reading and Reading Vocabulary, and smaller average losses in NCE scores in Reading Comprehension scores for ARI than for non-ARI schools.

NCEs and Longitudinal Effect Sizes for Cohort A and Non-ARI Schools

		Cohort A Schools				Non-ARI Schools			
Subtest	Grades	1998 Mean	2000 Mean	00-98 Diff	Effect Size	1998 Mean	2000 Mean	00-98 Diff	Effect Size
Total Reading	3-5	51.01	51.78	0.77	0.04	49.78	51.57	1.79	0.09
Reading Vocabulary	3-5	52.11	53.54	1.43	0.08	51.31	52.56	1.25	0.06
Reading Comprehension	3-5	50.99	50.83	-0.16	-0.01	49.63	50.83	1.20	0.06
Total Reading	6-8	56.40	59.34	2.94	0.16	52.48	52.10	-0.38	-0.02
Reading Vocabulary	6-8	54.72	56.05	1.33	0.08	51.88	49.99	-1.89	-0.11
Reading Comprehension	6-8	56.82	60.40	3.58	0.19	52.63	53.15	0.52	0.03
Total Reading	9-11	28.56	30.73	2.17	0.17	45.39	46.12	0.73	0.04
Reading Vocabulary	9-11	29.70	35.44	5.74	0.44	44.44	49.62	5.18	0.30
Reading Comprehension	9-11	31.12	30.38	-0.74	-0.06	46.95	44.69	-2.26	-0.13

Comparison of longitudinal scores for students in Cohort B favored students in ARI schools in all reading subtests and all elementary and middle school grades with the exception of Reading Vocabulary for students in Grade 7 in 1999. Differences between ARI and non-ARI secondary schools favored non-ARI schools on all reading subtests with the exception of Reading Vocabulary between Grade 9 and Grade 10 (1999-2000).

NCEs and Longitudinal Effect Sizes for Cohort B and Non-ARI Schools

		Cohort B Schools				Non-ARI Schools			
Subtest	Grades	1999 Mean	2000 Mean	00-99 Diff	Effect Size	1999 Mean	2000 Mean	00-99 Diff	Effect Size
Total Reading	3-4	48.43	53.32	4.89	0.25	49.88	53.26	3.38	0.17
Reading Vocabulary	3-4	50.11	53.93	3.82	0.19	51.28	53.86	2.58	0.13
Reading Comprehension	3-4	48.37	52.17	3.80	0.20	49.79	52.31	2.52	0.13
Total Reading	4-5	53.59	53.26	-0.33	-0.02	53.00	51.57	-1.43	-0.07
Reading Vocabulary	4-5	53.78	53.81	0.03	0.00	53.47	52.56	-0.91	-0.04
Reading Comprehension	4-5	52.78	52.46	-0.32	-0.02	52.17	50.83	-1.34	-0.07
Total Reading	6-7	53.95	57.10	3.15	0.16	52.69	50.37	-2.32	-0.12
Reading Vocabulary	6-7	53.20	53.79	0.59	0.03	52.59	48.16	-4.43	-0.26
Reading Comprehension	6-7	53.81	57.95	4.14	0.20	52.54	51.35	-1.19	-0.06
Total Reading	7-8	53.97	56.49	2.52	0.12	49.96	52.10	2.14	0.10
Reading Vocabulary	7-8	51.28	52.70	1.42	0.07	47.67	49.99	2.32	0.11
Reading Comprehension	7-8	54.91	58.21	3.30	0.16	51.02	53.15	2.13	0.10
Total Reading	9-10	52.14	51.62	-0.52	-0.03	45.59	45.07	-0.52	-0.03
Reading Vocabulary	9-10	50.56	56.58	6.02	0.36	44.86	50.45	5.59	0.32
Reading Comprehension	9-10	53.27	49.02	-4.25	-0.24	47.01	43.24	-3.77	-0.21
Total Reading	10-11	51.68	52.50	0.82	0.04	44.73	46.12	1.39	0.07

Reading Vocabulary	10-11	56.02	55.67	-0.35	-0.02	50.02	49.62	-0.40	-0.02
Reading Comprehension	10-11	49.19	50.46	1.27	0.08	42.95	44.69	1.74	0.09

QUESTION 2: WHICH ARI SCHOOLS ARE MAKING PROGRESS TOWARD 100% LITERACY AND WHICH ARE NOT?

Summary of Findings for Question 2

Findings: Overall

- Thirteen of the 15 Cohort A schools and 45 of the 61 Cohort B schools with *Stanford 9* demonstrated positive effect sizes (ranging from .01 to .72) across time on one or more of the reading subtests.

Findings: Struggling Readers

- Ten of the 15 Cohort A schools and 47 of the 61 Cohort B schools with *Stanford 9* data showed decreases in the percentage of students (.02% to 18%) scoring in stanines 1-3 on one or more of the reading subtests

Findings: Readers At or Above Grade Level:

- Thirteen of the 15 Cohort A schools and 51 of 61 Cohort B schools with *Stanford 9* data showed increases in the percentage of students (.05% to 28%) scoring at or above the fifth stanine on one or more of the reading subtests.

Findings: Collateral Outcomes

- Systematic and substantial decreases in discipline and special education referrals were found in the subset of ARI schools that reported such data. Discipline referrals in the reporting Group A schools (7 out of 16) decreased by 67%, from 1795 in 1998 to 596 in 2000. The average decrease in discipline referrals across the reporting Group B schools (27 out of 65) was 23%.
- As a group, the eight reporting Group A schools decreased special education referrals by 28% from 1998 to 2000. The 33 reporting Group B schools decreased special education referrals by 14% from 1999 to 2000.

Detailed Findings for Question 2

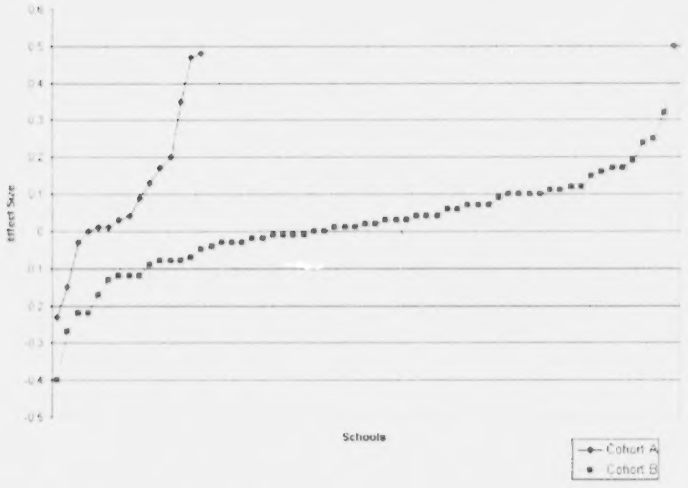
Findings: Overall

More than half of Cohort A and Cohort B schools demonstrated positive effect sizes across time on Total Reading, Reading Comprehension, and/or Reading Vocabulary subtests. Effect sizes for the 15 Cohort A schools and the 61 Cohort B schools ranged from moderately negative to moderately positive in each cohort for each reading subtest. These findings suggest considerable cross-sectional variability in the average *Stanford 9* performance of students in ARI schools.

Thirteen of the 15 Cohort A schools and 45 of the 61 Cohort B schools with *Stanford 9* data demonstrated positive effect sizes (ranging from .01 to .72) across time on one or more of the reading subtests. Seven Cohort A schools and 27 Cohort B schools showed positive effect sizes on all three reading subtests.

Among the 15 Cohort A schools included in the analysis, 11 showed gains in NCEs from 1998 to 2000 on Total Reading. The effect sizes for the 15 schools ranged from .-.23 to 48. Among the 61 Cohort B schools in the analysis, 34 of them showed NCE gains from 1999 to 2000 in Total Reading. The effect sizes for the 61 Cohort B schools ranged from -.40 to .50. The figure below illustrates the variability in effect sizes graphically. The shorter line depicts the 15 Cohort A schools ranging from the greatest loss (effect size = -.23) to the greatest gain (effect size = .48). This represents a range of almost three fourths of a standard deviation. Likewise, the longer line depicts the range of performances for the 61 Cohort B schools, varying from a loss of -.40 to a gain of .50. This represents a range of almost a full standard deviation. Figures similar to this one, showing the variability of each Cohort on each *Stanford 9* subtest, can be found in Appendix H.

Mean NCE Effect Sizes for Cohort A and Cohort B Schools on the Stanford 9 Total Reading Subtest



A total of 10 schools in Cohort A and 34 schools in Cohort B showed gains on Reading Vocabulary. The effect sizes ranged from -.15 to .72 for Cohort A and from -.34 to .42 for Cohort B.

A total of 9 schools in Cohort A and 37 schools in Cohort B showed gains on Reading Comprehension. The effect sizes ranged from -.29 to .51 for Cohort A and from -.37 to .51 for Cohort B.

Findings: Struggling Readers

Ten of the 15 Cohort A schools and 47 of the 61 Cohort B schools with *Stanford 9* data showed decreases in the percentage of students (.02% to 18%) scoring in stanines 1-3 on one or more of the reading subtests. Seven Cohort A schools and 29 Cohort B schools showed decreases on all three reading subtests.

Of the 15 Cohort A schools included in the analyses, 7 showed decreases from 1998 to 2000 in the percentage of struggling readers for Total Reading, 8 showed decreases for Reading Vocabulary, and 9 showed decreases for Reading Comprehension. Percentage change ranged from a 38% decrease to a 53% increase

for Total Reading, from a 44% decrease to a 110% increase for Reading Vocabulary, and from a 43% decrease to an 82% increase for Reading Comprehension.

Among the 61 Cohort B schools included in the analyses, 41 showed decreases from 1999 to 2000 in the percentage of struggling readers for Total Reading, 39 showed decreases for Reading Vocabulary, and 41 showed decreases for Reading Comprehension. Percentage change ranged from a 56% decrease to a 72% increase for Total Reading, from a 72% decrease to a 79% increase for Reading Vocabulary, and from a 56% decrease to an 85% increase for Reading Comprehension.

Findings: Readers At or Above Grade Level:

Thirteen of the 15 Cohort A schools and 51 of the 61 Cohort B schools with *Stanford 9* data showed increases in the percentage of students (.05% to 28%) scoring at or above the fifth stanine on one or more of the reading subtests.

Of the 15 Cohort A schools included in the analyses, 11 showed increases from 1998 to 2000 in the percentage of readers reading at or above grade level on Total Reading, 11 showed increases for Reading Vocabulary, and 12 showed increases for Reading Comprehension. Percentage change ranged from a 14% decrease to a 67% increase for Total Reading, from a 14% decrease to a 72% increase for Reading Vocabulary, and from a 14% decrease to a 65% increase for Reading Comprehension.

Among the 61 Cohort B schools included in the analyses, 32 showed increases from 1999 to 2000 in the percentage of students scoring at or above grade level on Total Reading, 38 showed increases for Reading Vocabulary, and 39 showed increases for Reading Comprehension. Percentage change ranged from a 23% decrease to a 53% increase for Total Reading, from an 18% decrease to a 34% increase for Reading Vocabulary, and from a 17% decrease to a 55% increase for Reading Comprehension.

Findings: Longitudinal Results

Nine **Cohort A** schools served students in Grades 3-5 between 1998 and 2000. The 1998 test results for students who were in Grade 3 are baseline for this group of students because they had not received ARI services at that point. By tracking groups of students through Grades 4 and 5, we are able to understand the impact of ARI on a common group of students across two years. This longitudinal data is presented in Appendix G.

Positive longitudinal effects (effect sizes greater than zero) were found in at least one subtest for eight of the nine Cohort A elementary schools (Grades 3 through 5) reporting appropriate data. Of the seven schools, six demonstrated improvements in Total Reading, six in Reading Vocabulary, and three in Reading Comprehension. The sole middle school with longitudinal data (grades 6 through 8) yielded increases in Total Reading and Reading Comprehension. The sole secondary school (grades nine through eleven) reported gains in Total Reading and Reading Vocabulary.

In addition to the performance gains in Total Reading and Reading Comprehension, several schools demonstrated large effects for other specific subtests. For example, one school demonstrated effect size gains in excess of a full standard deviation (ES > 1.0) for reading vocabulary, math total, social science, and total battery. Another school demonstrated an effect size gain of greater than one for social science. An effect size gain of 1.0 or greater for science and social science occurred in another school. In contrast, one school demonstrated effect size declines in excess of one standard deviation for six of seven subtests.

Positive longitudinal effects were found in at least one subtest for 25 of the 43 Cohort B elementary schools with necessary data for Grades 4 and 5. Of these 25 schools, 16 yielded gains in Total Reading, 20 in Reading Vocabulary, and 18 in Reading Comprehension. Nine of the 11 middle schools reporting performance for

Grades 7 and 8 in Cohort B yielded positive effect sizes for at least one of the reading subtests. All nine of the schools demonstrated gains in Reading Comprehension, seven of the nine showed gains in Total Reading, and six of the nine gained in Reading Vocabulary. Three of the four Cohort B secondary schools with longitudinal data for Grades 10 and 11 demonstrated gains on at least one of the reading subtests. Of these three schools, all three reported effect size gains in Reading Comprehension, two in Total Reading, and two in Reading Vocabulary.

Positive longitudinal effects were found in at least one subtest for 41of 46 Cohort B elementary schools with necessary data for Grades 3 and 4. Of these 41 schools, 39 yielded gains in Total Reading, 32 in Reading Vocabulary, and 39 in Reading Comprehension. Of the 46 elementary schools, 31 reported gains in all reading subtests between Grades 3 and 4. Nine of the 12 middle schools reporting performance for Grades 6 and 7, and eight middle schools in Cohort B, yielded positive effect sizes for at least one of the reading subtests. Of the nine schools, eight demonstrated gains in Reading Comprehension, five in Total Reading and two in Reading Vocabulary. Three of four Cohort B secondary schools with longitudinal data for Grades 9 and 10 demonstrated gains on at least one of the reading subtests between these grades. Of these three schools, all three reported effect size gains in Reading Vocabulary, two in Total Reading, and none in Reading Comprehension

Findings: Collateral Outcomes

In addition to academic outcomes assessed by the *Stanford 9*, this study explored the impact of ARI on collateral outcomes that are related to learning environments and learning opportunities for students. Three collateral outcomes were explored: discipline referrals, special education referrals, and library circulation. Although ARI did not require that principals maintain such data, principals were asked to provide data for the last three years (1997-1998, 1998-1999 and 1999-2000 for Cohort A

and 1998-1999 and 1999-2000 for Cohort B) on discipline referrals, special education referrals, and library circulation. Those findings are reported below, and complete information by school within each cohort can be found in Appendix I.

Discipline Referrals:

Among the Cohort A schools that began implementing ARI during 1998, seven principals reported three years of discipline referral data. Five principals reported a steady decrease in referrals. The other two principals reported mixed results, but yielded an overall decrease in referrals. From a baseline of 1,795 referrals across these seven schools, the final number of referrals in 2000 was 596. This is a decrease of 67%.

Four Cohort A schools reported two years of discipline referral data. One principal provided data from 1998 (first year of implementation) and 2000 and evidenced a 64% decrease in referrals. Although a substantial gain in referrals for one school occurred between 1999 and 2000, it is not possible to determine the degree of change in referrals since the beginning of ARI in this school or for another school that reported a 16% decrease from 1999 to 2000. One principal reported a 53% decrease from 1998 to 1999. Three principals reported only one year of data. Two principals reported no data.

Thirty-five principals in Cohort B schools reported two years of discipline referral data. The majority (27 of 35) of principals in schools that began ARI during 1999 reported a decrease in discipline referrals between 1999 and 2000. Of the six principals reporting increases, two offered explanations (new dress code/increased enrollment and new school opened). Two schools reported no change. The average decrease in discipline referrals across all reporting schools was 23%. Five principals reported only one year of data. Twenty-one principals reported no data, and principals from four schools did not respond to the survey.

Special Education Referrals:

Eight Cohort A principals reported three years of data concerning special education referrals. Two principals reported decreases each year, and four principals noted decreases from 1998 to 2000. One principal indicated an increase of only one student in 1999. Another principal reported a substantial increase from 1998 to 1999 and then a decrease between 1999 to 2000 with the net result of an overall increase in referrals from 1998 to 2000. As a group, the eight reporting schools yielded an overall decrease in special education referrals of 28% from 1998 to 2000.

Two Cohort A principals reported two years of data (1999 and 2000). Of these, one showed an increase in students being referred, and the other showed a decrease of three students being referred for special education. One principal reported only one year of data. Four principals did not provide data concerning special education referrals.

Thirty-three Cohort B principals reported data for both 1999 and 2000. Twenty principals noted a decrease in special education referrals, three principals indicated no change, and ten principals reported an increase in referrals. The overall change in referrals for these 33 schools was a decrease of 14%. Three principals reported only one year of special education referral data. Twenty-five principals provided no data, and principals from four schools did not respond to the survey.

Library Circulation:

Five of seven Cohort A principals who recorded three years of data for library circulation reported increases in circulation. One principal noted decreases in circulation each year. One principal reported an increase followed by a decrease, which was explained as being due to the school's restructuring (from K-6 to K-3). The overall increase in library circulation for the seven schools with three years of data was 44% between 1998 and 2000.

Five principals from Cohort A reported two years of library circulation data (1999-2000). All principals reported an increase in library circulation during this period. Two principals provided only one year of data. Two principals reported no data.

Principals in forty schools in Cohort B reported two years of library circulation data. Of these principals, 37 reported increases in library circulation. Only three principals recorded declines in circulation. The overall increase in library circulation was 40%. Seven principals reported only one year of data. Fourteen principals provided no data, and principals from four schools did not respond to the survey.

Findings: Identification of Higher and Lower Performing Schools

Student outcomes varied from school to school. Some schools demonstrated marked improvement in performance on the *Stanford 9*, whereas others showed losses, and some had mixed results. In order to respond to the third and fourth evaluation questions concerning links, subsets of higher and lower performing schools were identified using *Stanford 9* results. Cross-sectional data for all eight subtests (Reading Comprehension, Reading Vocabulary, Total Reading, Total Math, Total Language, Science, Social Science, and Total Battery) and all three analyses (percentage of students in stanines 1-3, percentage of students in stanine 5 and above, and NCE effect sizes) as well as NCE effect sizes for longitudinal results were used to compile a body of evidence on which to base the grouping of schools into higher and lower performing categories. **Essentially, higher-performing schools are those with good cross-sectional (effect sizes of .10 or greater) or longitudinal performance (effect sizes of .20 or greater) on Reading Comprehension or Total Reading NCE scores and a trend toward positive outcomes across a number of other *Stanford 9* performance indicators (average NCE scores and changes in percentages of students in the lowest three stanines and top five stanines across *Stanford 9* subtests). Lower performing schools were those with poor cross-sectional (effect sizes of -.10 or**

less) or poor longitudinal performance (effect sizes of -.20 or less) on Reading Comprehension or Total Reading and a trend toward negative outcomes across a number of other *Stanford 9* performance indicators. The specific procedures used for identifying higher and lower performing schools involved several steps or stages as described below.

1. The first step in the process involved gathering evidence using cross-sectional results.
    a. The NCE effect sizes for cross-sectional data were examined for schools in each cohort. Schools were initially selected as higher performing if they had effect sizes of more than .10 on either Total Reading or Reading Comprehension. They were initially selected as lower performing if they had effect sizes of less than -.10 on either Total Reading or Reading Comprehension.
    b. Effect sizes within each of these subsets of schools were then examined for the other six subtests to see if, on these tests, the same criteria (more than .10 or less than -.10) (1) were met, (2) were consistent (in the same direction) but not to the .10 or -.10 levels, or (3) were inconsistent (in the opposite direction) with the Total Reading or Reading Comprehension findings.
    c. For these same subsets of schools (identified in 1.a above) a similar analysis was done using changes in the percentage of students scoring in stanines 1-3 for each subtest. A decrease of more than 5% was the benchmark for meeting the criterion for the higher performing group, and an increase of more than 5% was the benchmark for meeting the criterion for the lower performing group. Differences in the percentage of students scoring in stanines 1-3 were examined for all eight subtests to see if the

criteria (1) were met, (2) were consistent (in the same direction) but not to targeted levels of 5% increase or 5% decrease, or (3) were inconsistent (in the opposite direction) of the Reading Comprehension or Reading Total NCE performance that placed them in the initial group.

d. Finally, for the same subset of schools identified in 1.a above, an analysis also was done using the change in percentage of students scoring at or above the 5th stanine for each *Stanford 9* subtest. An increase of more than 5% was the benchmark for meeting the criterion for the higher performing group, and a decrease of more than 5% was the benchmark for meeting the criterion for the lower performing group. Differences in the percentage of students scoring at or above the 5th stanine were examined for all eight subtests to see if the criteria (1) were met, (2) were consistent (in the same direction), or (3) were inconsistent (in the opposite direction).

e. Using the above three sets of information (1.b, 1.c, and 1.d), 24 possible points of analysis were considered when examining the initial membership of schools in higher or lower performing groups (1.a): NCE effect sizes for 8 subtests, differences in the percentage of students scoring in stanines 1-3 for 8 subtests, and differences in the percentage of students scoring at or above stanine 5 for 8 subtests. **If a school met the criteria for inclusion in the initial group (1.a above) and met or was consistent on at least 12 of the 24 points of analysis, it was selected for tentative inclusion in the higher or lower performing group.** These classified schools would be examined one more time for the appropriateness of their fit using additional *Stanford 9* data discussed in the next section. Of the 15 Cohort A schools with *Stanford 9* data for

any grade that included Grades 3 through 11, five schools were tentatively identified as higher performing schools and three were tentatively identified as lower performing schools. Of the 61 Cohort B schools with *Stanford 9* data, fourteen were tentatively identified as higher performing schools, and seven were tentatively identified as lower performing schools.

2. The next phase involved the examination of longitudinal results for those schools with such data available.

    a. **Longitudinal effect sizes for all 8 subtests were examined for those schools already tentatively identified in step 1 above**. If half or more of the longitudinal effect sizes were consistent with the direction of the cross-sectional effect sizes (that is, positive for the higher performing group and negative for the lower performing group), schools remained in the group. This step resulted in the removal of one school from the Cohort A higher performing group, two schools from the Cohort A lower performing group, and three schools from the Cohort B higher performing group.

    b. Longitudinal effect sizes also were examined for schools that had not yet been placed in a group to see if longitudinal results might identify any additional schools. Schools were tentatively added to the list of higher performing schools if their longitudinal effect sizes were .20 or greater on either the Total Reading or the Reading Comprehension subtests. Schools were tentatively added to the list of lower performing schools if their longitudinal effect sizes were -.20 or less on either the Total Reading or the Reading Comprehension subtests.

c. Using the sets of schools identified in 2.b above, longitudinal effect sizes for the other six subtests and cross-sectional effect sizes for all eight subtests were examined to see if the criteria (1) were met (greater than .20 or less than -.20), (2) were consistent (in the same direction) but not to the criterion level, or (3) were inconsistent (in the opposite direction) from the Total Reading or Reading Comprehension score that placed the school in the group. Schools remained on the list if at least half the longitudinal effect sizes met the criteria or were consistent and at least half the cross-sectional effect sizes either met the criteria or were consistent. This step resulted in the addition of two schools to the Cohort A higher performing group, one school to the Cohort A lower performing group, and three schools to the Cohort B higher performing group.

3. Finally, for those schools without *Stanford 9* data for Grades 3 through 11, additional *Stanford 9* Grade 1 and Grade 2 data, if available from the schools, were examined to determine if evidence existed to include the school in either the higher or lower performing group. One such Cohort A school provided Grade 2 data. These data indicated a two-year increase of 8 percentile points in Reading Comprehension and 6 percentile points in Total Reading. This school was added to the Cohort A higher performing group. One such Cohort B school provided Grade 2 data. Results did not support inclusion in either the higher or lower performing group.

Results of these steps were summarized for each school selected. Those schools selected for each group are identified in the tables below, along with the results of the analysis of the criteria. Seven higher performing and two lower performing schools were identified in Cohort A, and fourteen higher performing schools and eight

lower performing schools were identified in Cohort B. Other schools may have demonstrated gains (or losses) in particular areas but are not included in the groupings. This is because the intent was to base grouping decisions on a preponderance of evidence rather than a single piece of evidence. The groupings form the basis for later discussions regarding factors related to student outcomes. **It should be noted that the criteria for identifying higher and lower performing schools were chosen by the evaluators. Different sets of criteria might result in the selection of different groups of schools.**

ARI Cohort A – Higher Performing Schools

	Cross-sectional Analyses			Longitudinal Analyses				
School	Number of Criteria Met	Number of Consistent Findings	Number of Inconsiste nt Findings	Longi-tudinal Grades	Total Reading Effect Size	Reading Comprehension Effect Size	Effect Size Range (All 8 Subtests)	Number of Negative Effect Sizes
48	22	2	0	3-5	.56	-.06	-.06 to 2.03	1
24	22	2	0	None				
15	24	0	0	9-11	.17	-.06	-.43 to .44	3
80	18	6	0	3-5	.36	.42	-.13 to .67	1
74*	4	9	11	3-5	.64	.51	.31 to 1.29	0
20*	1	11	12	6-8	.17	.25	.00 to .25	0
04	*Stanford 9* Grade 2 showed percentile gains of 5 to 8 percentile points.							

*First identified using longitudinal results.

ARI Cohort A – Lower Performing Schools

	Cross-sectional Analyses			Longitudinal Analyses				
School	Number of Criteria Met	Number of Consistent Findings	Number of Inconsiste nt Findings	Longi-tudinal Grades	Total Reading Effect Size	Reading Comprehension Effect Size	Effect Size Range (All 8 Subtests)	Number of Negative Effect Sizes
14	14	7	3	3-5	.02	-.04	-.18 to .21	2
78*	11	5	8	3-5	-1.66	-1.26	-.57 to -2.01	8

*First identified using longitudinal results.

ARI Cohort B – Higher Performing Schools

	Cross-sectional Analyses			Longitudinal Analyses				
School	Number of Criteria Met	Number of Consistent Findings	Number of Inconsiste nt Findings	Longi-tudinal Grades	Total Reading Effect Size	Reading Comprehension Effect Size	Effect Size Range (All 8 Subtests)	Number of Negative Effect Sizes
36	17	1	6	3-5	1.25	.98	.35 to 1.57	0
47	24	0	0	3-5	-.06	-.11	-.11 to .26	3
05	20	4	0	6-8	.34	.39	.05 to .39	0
02	17	7	0	None				
46	10	12	2	None				
71	10	13	1	3-5	.17	.17	.00 to .17	0
52	19	4	1	3-5	.84	.64	.64 to 1.40	0
29	12	12	0	3-5	.00	.05	-.23 to .10	3
53	21	2	1	3-5	.00	-.10	-.10 to .24	1
21	6	17	1	6-8	.35	.32	.00 to .49	0
65	18	6	0	3-5	.06	.07	-.12 to .27	1
28*	0	10	14	3-5	.54	.52	-.13 to .58	1
44*	4	11	9	3-5	.20	.14	.14 to .37	0
57*	5	9	10	3-5	.22	-.06	-.10 to 1.00	2

*First identified using longitudinal results.

ARI Cohort B – Lower Performing Schools

	Cross-sectional Analyses			Longitudinal Analyses				
School	Number of Criteria Met	Number of Consistent Findings	Number of Inconsiste nt Findings	Longi-tudinal Grades	Total Reading Effect Size	Reading Comprehension Effect Size	Effect Size Range (All 8 Subtests)	Number of Negative Effect Sizes
17	4	15	5	3-5	-.54	-.52	-.62 to -.32	8
76	15	8	1	3-5	-.08	.13	-.28 to .28	6
32	3	18	3	6-8	-.15	-.11	-.22 to .08	7
64	5	12	7	None				
12	15	8	1	3-5	-.30	-.28	-.42 to -.24	8
38	20	3	1	3-5	-.88	-.84	-1.35 to -.35	8
56	23	1	0	3-5	-.29	-.29	-.37 to -.11	8
45	24	0	0	3-5	-.73	-.61	-1.01 to -.61	8

Following identification of higher and lower performing schools, *Stanford 9* results for all schools in each group were analyzed. Analyses included the percentage of students scoring in stanines 1 through 3, the percentage of students scoring at or above stanine 5, and NCE means and effect sizes. Results for Total Reading, Reading Vocabulary, and Reading Comprehension are summarized in the tables below. Results for all eight subtests can be found in Appendix F. Case studies that were conducted to explain negative change apparent in two low-performing schools can be found in Appendix E.

Percentage Change: Students Scoring in Stanines 1-3

Subtest	Percentage Change			
	Cohort A High	Cohort A Low	Cohort B High	Cohort B Low
Total Reading	-13.35	20.40	-24.83	29.19
Reading Vocabulary	-7.01	30.39	-24.06	27.90
Reading Comprehension	-12.26	9.43	-20.38	29.90

Percentage Change: Students Scoring at or above Stanine 5

Subtest	Percentage Change			
	Cohort A High	Cohort A Low	Cohort B High	Cohort B Low
Total Reading	2.15	-3.01	11.56	-6.54
Reading Vocabulary	5.46	-9.51	10.85	-5.51
Reading Comprehension	3.29	4.16	10.29	-6.74

Mean NCE Difference Effect Sizes

Subtest	Effect Sizes			
	Cohort A High	Cohort A Low	Cohort B High	Cohort B Low
Total Reading	0.06	-0.02	0.16	-0.16
Reading Vocabulary	0.10	-0.13	0.18	-0.15
Reading Comprehension	0.04	0.04	0.13	-0.15

Clearly, differences exist between higher and lower performing groups. Although Reading Comprehension, Reading Vocabulary, and Total Reading are highlighted here, they were not the only subtests used to identify higher and lower performing schools, as noted earlier. Other subtests included Total Math, Total Language, Science, Social Science, and Total Battery. Whereas higher performing schools performed better than lower performing schools across all three reading subtests, the most obvious difference between higher and lower performing ARI schools was the schools' ability to reduce the percentage of struggling readers in the school. This reflects ARI efforts to improve reading skills among struggling readers (students scoring in stanines 1-3).

QUESTION 3:	WHAT FACTORS ARE RELATED TO SCHOOL OUTCOMES?

Summary of Findings for Question 3

Findings: Teachers

- Higher ARI implementation rates occurred in higher performing schools than in lower performing schools as determined by teacher survey responses.
- More hours of professional development occurred during the school year in higher performing schools than in lower performing schools as reported by teacher survey responses.
- Teachers responding to survey items reported that ARI had positive impacts on their teaching and student learning. Changes in teaching included use of current strategies, increases in time dedicated to student reading, increased confidence and enthusiasm regarding reading instruction, increased awareness of struggling readers, and positive school-wide change in the learning environment. Changes reported in student learning included improved reading skill, improved writing abilities, and increased motivation to read.

Findings: Reading Specialist

- Principals and teachers in higher performing schools rated the helpfulness of reading specialists higher than did principals and teachers in lower performing schools.
- Teachers in both higher and lower performing schools in Cohort A and Cohort B considered the role of the reading specialist as important to the successful implementation of ARI.

Findings: Principal

- Principals in higher performing schools were perceived by teachers and higher education partners as taking a greater leadership role in implementing ARI than their counterparts in lower performing schools.
- Correlations between principal leadership scores as evaluated by the teachers, higher education partners, and reading specialists and *Stanford 9* reading score effect sizes for Cohort B were .20 or greater.

Findings: Higher Education Partner

- A greater proportion of teachers in higher performing schools than in lower performing schools indicated that the higher education partner was valued as a model for instruction and as a resource for professional development.

Identification of Mediating Factors

A number of factors that might mediate student outcomes were initially considered as part of this evaluation. They included the following:

- Teacher participation in Summer Institutes
- Principal participation in Summer Institutes
- Ongoing professional development
- Teacher implementation of ARI
- Reading specialist activities and helpfulness
- Principal leadership in implementing ARI
- Implementation of the intervention plan
- Higher education partner activities and helpfulness
- Teacher turnover
- Principal turnover

Following the collection of data, four of these factors were not included in the final analysis. The four factors and their reasons for exclusion are indicated below.

- Teacher participation in Summer Institutes – ARI requires its schools to have 85% of its faculty participate in the Summer Institute.
- Principal participation in Summer Institutes – ARI requires that principals attend the Summer Institute.
- Teacher turnover – Timing of data collection and the manner in which the question was asked resulted in data that were not useful in responding to this factor.
- Principal turnover - Timing of data collection and the manner in which the question was asked resulted in data that were not useful in responding to this factor.

Two primary quantitative procedures are available for identifying the impact of ARI implementation factors on achievement outcomes. The first approach involves comparisons of means and/or frequencies and percentages between groups. For the purpose of this study, schools were grouped into one of two levels of academic achievement – higher performing and lower performing schools. Descriptions of the protocol used to classify schools are provided earlier in the report. Means (or frequencies and percentages) of potentially mediating factors between higher and lower performing schools are presented and compared. Variables that reveal substantial differences for higher performing and lower performing school clusters appear to have some degree of impact on the category to which schools are assigned.

The second approach to identifying variables that mediate the effects of ARI on student achievement outcomes involves correlating scores for the mediating variable with outcome scores. Correlations range in value between -1.0 and +1.0. Negative

numbers suggest that increases in one score are associated with decreases in the other score. Positive correlations indicate that increases in one score are associated with increases in the other score. For example, to the extent that mediating factors such as principal leadership scores are positively correlated with outcomes such as *Stanford 9* effect size gains, we can conclude that increases in the mediator are associated with gains in the outcome. This does not imply that gains in the mediator cause gains in the outcome. Correlations do not guarantee causation.

Survey Response Rates

Surveys were the primary source of data collected on possible factors related to school outcomes. Surveys were distributed at meetings held in September with principals and/or their representatives from the 81 ARI schools. The following table indicates the number of schools represented and the number of surveys received for each category of respondent.

Survey Responses

Type of Survey	Number of Surveys	Number of Schools Represented	Number of Cohort A Surveys	Number of Cohort B Surveys
Teacher	1860	81	360	1500
Principal	77	77	16	61
Reading Specialist	74	74	14	60
Higher Education Partner	44	43	9	35

These surveys represent a 95% response rate for principals, a 91% response rate for reading specialists, and a 54% response rate for higher education partners. Using principal reports of the number of returning teachers in the schools, the overall teacher response rate was found to be 86% for the 75 schools with valid teacher population data from principals. (The response rate is based on 1,717 teachers responding from these schools out of 1,990 teachers reported by the principal to have been returning teachers in 2000-2001.) The response rate varied considerably from school to school. Among these 75 schools, two schools had a response rate of less than 50%; 19 schools had a 51% to 75% response rate; another 19 schools had 76% to

85% response rate; and 35 schools had a response rate of 86% or better. Complete information on response rates by school can be found in Appendix I. In a few cases the teacher response rates shown exceed 100%. This may be explained by incomplete or inaccurate information reported on the principal survey.

Following are the survey response rates from higher education faculty. Content of responses is detailed in the section on Pre-service Component Results.

Higher Education Reading Faculty Survey Responses

	Number of Surveys	Number of Institutions Represented	Number from Public Institutions	Number from Private Institutions	Number Unknown
Higher Education Faculty	49	21	35	12	2

Rating Scale Characteristics

The various surveys contain rating scales asking respondents to rate various aspects of ARI. These included activities of the reading specialist, higher education partner, and principal as well as a scale to rate teacher implementation of ARI and the implementation of the intervention plan at the LDS. The rating scale items were examined individually as well as summed to achieve a total score for the scale. Internal consistency reliabilities of the scales were assessed prior to analyzing summed scales. Cronbach alphas for the scales ranged from .83 to .96 with 13 of the 18 scales achieving an alpha coefficient of .90 or better. (High internal consistency reliabilities mean that response patterns for an individual tend to be consistent across items. This is often interpreted to mean that the items seem to be measuring the same construct.) The table below delineates the number of items, the response scale, and the minimum and maximum number of points possible on each scale. In cases where a response was absent, it was given a value equal to the lowest value of the scale on which it was found.

Rating Scale Characteristics

Survey	Number of Items	Response Scale	Minimum Possible	Maximum Possible
Reading Specialist Frequency	14	1 to 3 (never, occasionally, frequently)	14	42
Reading Specialist Helpfulness	14	1 to 3 (not at all, somewhat, very)	14	42
Higher Ed Partner Frequency	14	1 to 3 (never, occasionally, frequently)	14	42
Higher Ed Partner Helpfulness	14	1 to 3 (not at all, somewhat, very)	14	42
Principal Leadership	17	1 to 4 (strongly disagree to strongly agree)	17	68
Teacher Implementation of ARI	7	0 to 3 (none, low, moderate, high)	0	21
Implementation of Intervention Plan	14	0 to 3 (none, low, moderate, high)	0	42

Detailed Findings for Question 3

Findings: Teachers

Teacher Reports of Impact of ARI:

In addition to academic and collateral quantitative results, this study also examined teachers' perceptions of impacts of ARI on their teaching and student learning. Teachers' responses to key questions are summarized in Appendix K. The following section summarizes teacher responses to these categories.

Changes in Teaching: In response to the survey question asking how teaching had changed as a result of the Summer Institute and/or implementing ARI, by far, teachers reported the greatest change in their teaching as increased attention to the use of strategies and reading instruction in their classrooms. Teachers in Cohort A higher performing schools more frequently reported an increase in student reading time and reading involvement than for the lower performing schools. Other categories of responses included greater attention to strategies and activities related to reading instruction, greater confidence or enthusiasm regarding reading and reading instruction, more assessment of reading, and an increased awareness of struggling readers or the needs of readers.

Changes in Student Learning: Changes in Student Learning: When asked how student learning changed in their classroom as a result of ARI training, teachers most

frequently cited improved reading or writing performance and an improved motivation or willingness to read. Many of the comments, particularly in several of the lower performing schools, focused on changes in teaching behaviors and are, therefore, not considered in this categorization. Teachers identified the biggest impact of ARI as improvement in student achievement in literacy and student motivation/engagement relative to literacy.

Ongoing Professional Development:

Teachers reported the number of hours of professional development they received in each module during the 1999-2000 school year. (For most teachers, this was following their participation in the Summer Institute.) Overall, approximately 40% to 49% of teachers indicated receiving some professional development. The largest percentage (26% to 32%) of this group received one to five hours on a module.

Patterns did differentiate lower and higher performing schools in Cohort B. A greater percentage of teachers in higher performing schools reported receiving some professional development. Furthermore, teachers in higher performing schools tended to receive more hours of professional development. Among the lower performing schools 33% to 43% received some professional development, and 7% to 13% received more than five hours. The higher performing schools reported 36% to 51% receiving some professional development with 16% to 36% receiving more than five hours. Complete results can be found in Appendix I.

Teacher Implementation of ARI:

Teachers were asked to rate their own implementation of each of the seven ARI modules on a scale ranging from 0 (no implementation) to 3 (high implementation). Teacher self-ratings for each module were summed to obtain a total score that could range from 7 to 21. School means for the summed scores ranged from 7.39 to 20.59, with an overall mean of 13.17. The table below provides total score means for teacher

implementation among higher and lower performing schools in each cohort. Teachers in the higher performing schools both in Cohort A and Cohort B tended to rate themselves higher in implementing ARI than did teachers in lower performing schools. (NOTE: Caution should be exercised in interpreting results involving Cohort A lower performing schools as this group includes only two schools.)

Scale Means for Teacher Self-Reports of Implementation of ARI

	Cohort A			Cohort B		
Survey	Higher Performers	Lower Performers	Difference	Higher Performers	Lower Performers	Difference
Teacher Implementation of ARI	13.50	10.13	3.37	14.92	13.19	1.73

The following table provides teacher implementation means among higher and lower performing schools in each cohort for each module. Overall, implementation level varied among the modules, but generally there was moderate (2.0 or greater) implementation for most modules. Higher performing schools in both Cohort A and Cohort B yielded higher implementation scores for each module than were found for lower performing schools. The largest differences between higher and lower performing schools were for Phonics/Phonemic Awareness (difference = .71) and Oral Language/Vocabulary Development (difference = .72) in Cohort A.

Means for Teacher Self-Reports of Implementation of ARI Modules

	Cohort A			Cohort B		
Module	Higher Performers	Lower Performers	Difference	Higher Performers	Lower Performers	Difference
Formal Assessment	1.16	1.13	.03	1.74	1.32	.42
Informal Assessment	1.90	1.55	.35	2.11	1.80	.31
Phonics/Phonemic Awareness	2.03	1.32	.71	2.09	1.89	.20
Comprehension Strategies	2.23	1.58	.65	2.37	2.20	.17
Oral Language/Vocabulary Development	2.27	1.55	.72	2.34	2.18	.16
Reading/Writing Connection	2.04	1.74	.30	2.15	1.99	.16
Effective Intervention	1.88	1.26	.62	2.13	1.80	.33

Teacher Implementation total scores were correlated with effect sizes of each of the *Stanford 9* subtests. All correlations were less than .10. Data regarding teacher

implementation also was gathered from principals and reading specialists, who agreed with teachers that the Oral Language/Vocabulary and Comprehension were the modules most frequently implemented by teachers. Similar agreement was also found concerning lower levels of implementation of the formal assessment and phonics/phonemic awareness modules. It should be noted that the phonics/phonemic awareness module is most applicable in grades K-2 and for teachers of struggling readers.

Teacher Needs: Teachers indicated that continued substantive, instructional, and personal support are needed. They specifically mentioned continuing or supplemental reading specialist support, additional training, follow-up communication and positive support from ARI staff, additional materials and resources, and changes in class size and curriculum.

Findings: Reading Specialist

Reading Specialist Activities and Helpfulness:

Principals, reading specialists, and teachers selected the frequency of each of 14 reading specialist activities using a scale from 1(never) to 3 (frequently). In addition, principals and teachers indicated the helpfulness of those activities on a scale from 1 (never) to 3 (frequently). The table below indicates mean responses among the higher and lower performing groups for Cohort A and Cohort B. The number of reading specialist and principal data points in the lower performing group in Cohort A consist of only one or two respondents. Differences were not calculated since they may not be meaningful.

Results relative to the average frequency ratings of activities by reading specialists were mixed. These average ratings were higher in higher performing schools in Cohort A and higher in the lower performing schools in Cohort B. The average

frequency ratings for reading specialist activities provided by teachers were lower than the ratings provided by principals and reading specialists.

Principals and teachers rated the helpfulness of the reading specialist in implementing various activities. Principals and teachers in higher performing schools rated the helpfulness of reading specialists higher than did the principal and teachers in lower performing schools. As with ratings of frequency of reading specialist activities, teachers' average ratings of the helpfulness of the reading specialist were lower than ratings by the principal.

Examination of correlations between *Stanford 9* effect sizes (standardized differences in NCE scores) and ratings of frequency of reading specialist activities by teachers, principals, and the reading specialist yielded no correlations of greater than .13. This suggests that there is no significant association between the frequency of such activities and *Stanford 9* performance for all subtests.

Examination of correlations between *Stanford 9* effect sizes and ratings of helpfulness of reading specialist activities as assigned by teachers and principals revealed no significant associations for the ratings by the principal. However, correlations in excess of .20 were found between reading specialist helpfulness ratings assigned by teachers with effect sizes from Reading Comprehension and Total Reading. This finding is consistent with teacher perceptions of greater helpfulness of the reading specialist in higher performing schools than in lower performing schools. Thus, both the correlations and teacher reports support greater helpfulness of the reading specialist in higher performing schools.

Scale Means for Ratings of Reading Specialist Frequency of Activities and Helpfulness

		Cohort A			Cohort B		
Survey	Rater	Higher Performers	Lower Performers	Difference	Higher Performers	Lower Performers	Difference
Reading Specialist Frequency of Activities	Principal	33.00	22.50		31.73	33.00	-1.27
Reading Specialist Frequency of Activities	Reading Specialist	32.50	32.00		31.07	33.71	-2.64
Reading Specialist Frequency of Activities	Teachers	27.33	22.45	4.88	29.99	28.63	1.36
Reading Specialist Helpfulness	Principal	32.86	18.50		32.60	32.14	.46
Reading Specialist Helpfulness	Teachers	26.00	21.81	4.19	28.79	26.15	2.64

Comments Concerning the Reading Specialist:

To better understand the perceived helpfulness of reading specialists, responses of teachers and principals to open-ended questions were summarized and coded. A discussion of these findings follows. Responses are coded into four major categories: 1) impact of the reading specialist on the reading abilities of struggling readers, 2) reading specialists' support of ARI implementation, 3) importance of having a reading specialist, and 4) additional comments.

Impact on Reading Abilities of Struggling Readers: Teachers identified a number of important ways in which the reading specialist impacted the reading abilities of struggling readers. The classification of responses into the following six categories did not clearly discriminate between higher performing and lower performing schools: worked with struggling readers directly, guided teachers and provided consultation, improved student reading, provided materials and resources, and served in an administrative/program supervision role.

The most frequent response from teachers representing higher and lower performing schools indicated that the reading specialist worked directly with struggling readers. The percentage of teachers citing this impact was somewhat greater in Cohort A than in Cohort B. More than 10% of the teachers in higher and lower performing

Cohort B schools indicated that the reading specialist provided a supervisory/administrative role, whereas only a few Cohort A teachers made this observation. Very few teachers (less than 10%) from both higher and lower performing schools indicated that the reading specialist did not substantially impact the reading ability of struggling readers in their schools.

Comments from principals supported teacher observations that the vast majority of reading specialists engaged in direct intervention with struggling readers. This was the common thread cited by the principals in both higher and lower performing schools (from higher performing schools, five of seven principals in Cohort A and nine of 13 principals in Cohort B; from lower performing schools, one of two principals in Cohort A and six of seven in Cohort B).

Reading Specialists' Support of ARI Implementation: When asked how the reading specialist supported the implementation of ARI, the same categories cited above emerged. Teachers identified "encouraging and assisting teachers" most often in both higher and lower performing schools, and the second most frequently noted category was "direct involvement with students."

In identifying the ways that the reading specialist supported ARI implementation, principals noted the reading specialists' modeling of teaching strategies, a factor that may be seen as differentiating higher performing schools from lower performing schools. This factor was indicated by six of the 20 principals from higher performing schools and one of eight principals from lower performing schools.

Importance of Having a Reading Specialist: Responses from **teachers** in all higher and lower performing schools were overwhelmingly positive when asked how important it was to have a reading specialist in their school. That is, regardless of the relative success of schools on standardized test outcomes or the duration of involvement

in ARI, at least 70% of teachers considered the reading specialist to be important to serving the needs of the students and teachers.

When asked to respond to the importance of having a reading specialist, **principals** were overwhelmingly positive. Nine of twenty principals from higher performing schools and one of eight principals in lower performing schools offered unqualified endorsement of the role of the reading specialist. Most of the other responses were either cautionary or negative concerning the pitfalls of the reading specialist role and the need for full time support. Higher performing principals noted that the reading specialist is overworked, needs to be qualified to work with teachers and be knowledgeable, that it is hard for faculty to look at a peer in a "supervisory" role, that it was difficult for the specialist to put aside personal feelings regarding strategies chosen by some of the grade levels. They also indicated that "more time should have been spent on coaching, modeling and workshops, and the time allotted for this was not used productively." Voiced repeatedly was "We need ONE per school."

Several comments from principals in lower performing schools focused on the need for full time help: "I wish we had a full time reading specialist at our school this year," and "For this program to be successful, I feel each school should have a full time specialist," and "The reading specialist should not be stretched so far (have so many other schools) that she cannot be effective. She should not be used as another teacher unit. She should have the FULL cooperation of the administration. I have great faith in this program and believe that it should not grow so fast that the reading specialist is lost in the shuffle or is not highly qualified for the job."

In addition, the principals from lower performing schools noted that "The reading specialists should spend the majority of his/her time assisting the teachers with strategies to enhance reading skills and/or modeling for teachers to observe. A reading specialist should not only identify problem areas, but also offer positive suggestions and

solutions to eliminate these problems," and "We need additional professional development for the specialist and staff members."

Another detrimental effect cited by principals was the inability of the reading specialist to relate effectively in a leadership role with the teaching staff. Three of the higher performing schools advised caution in the selection and expectations of the reading specialist. More importantly, without considering the schools that used a part-time resource, only two of the 20 higher performing schools (10%) expressed dissatisfaction with the work of their own reading specialist, while two of the eight lower performing schools (25%) expressed the same dissatisfaction. Therefore, only six of the 20 higher performing schools (30%) were hindered by these two factors, whereas seven of the eight lower performing schools (87.5%) had at least one of these obstacles to overcome. And where there were no obstacles, the process worked best. Perhaps the following comment from a higher performing school principal says it best, "You said the principal was the key. If so, the reading specialist is the ignition."

In general, teachers considered the role of the reading specialist as important to the successful implementation of ARI. This finding did not vary as a function of the Cohort or the relative success of the school. Furthermore, the role served by the reading specialist, as perceived by the teachers of the school, did not discriminate between higher and lower performing schools. It is noteworthy that a more administrative role emerged for some of the Cohort B reading specialists. When negative comments were made, such comments tended to indicate that teachers needed increased support from the specialist, increased affirmation of teachers' accomplishments and efforts, and a clear understanding of the role of the specialist.

Findings: Principal

Reading specialists, higher education partners, and teachers were all asked to provide information concerning the leadership of the principal in implementing ARI using

a scale from 4 (strongly agree) to 1 (strongly disagree). The following table indicates mean responses among the higher and lower performing groups for Cohort A and Cohort B.

Scale Means for Ratings of Principal Leadership

		Cohort A			Cohort B		
Survey	Rater	Higher Performers	Lower Performers	Difference	Higher Performers	Lower Performers	Difference
Principal Leadership	Reading Specialist	60.17	45.00		54.57	54.86	-.29
Principal Leadership	Higher Ed Partner	63.40	*		55.00	39.25	15.75
Principal Leadership	Teachers	51.06	37.13	13.93	51.86	44.64	7.22

*No responses were received in this group from higher education partners.

Principal leadership scores, as evaluated by teachers and higher education partners, discriminated higher performing and lower performing schools. Principals in higher performing schools were perceived as taking a greater leadership role in implementing ARI than were their colleagues in lower performing schools. Examination of correlations between principal leadership scores and *Stanford 9* effect sizes for Cohort B revealed correlations of .20 or greater between leadership and one or more of the reading scores, regardless of the source of the ratings. The correlation between the Reading Comprehension effect size and principal leadership, as evaluated by the reading specialist, was .20. The correlation between the Total Reading effect size and principal leadership, as evaluated by the higher education partner, was .20. The correlation between the Reading Comprehension effect size and principal leadership, as evaluated by the teacher, was .24. The correlation between the Total Reading effect size and principal leadership, as evaluated by the teacher, was .25. These findings highlight the importance of principal leadership in promoting academic gains in ARI schools.

Implementation of Intervention Plan at the LDS:

Principals and reading specialists were asked to indicate the level implementation of the school intervention plan using a scale of 0 (no implementation) to 3 (high implementation) on each of 14 elements of the plan. The following table indicates mean responses among the higher and lower performing groups for Cohort A and Cohort B. The number of principal and reading specialist data points among the lower performing schools in Cohort A consisted of only one or two respondents. Differences were not calculated since they may not be meaningful.

Scale Means for Ratings of Implementation of the Intervention Plan

		Cohort A			Cohort B		
Survey	Rater	Higher Performers	Lower Performers	Difference	Higher Performers	Lower Performers	Difference
Implementation of Intervention Plan	Principal	31.85	27.00	NA	31.57	36.57	-5.00
Implementation of Intervention Plan	Reading Specialist	33.92	25.56	NA	29.78	31.50	-1.72

A surprising effect regarding implementation of the intervention plan was that both principals and reading specialists in the lower performing schools of Cohort B reported higher average implementation rates than their colleagues in higher performing schools. This is in contrast to the opened-ended survey responses where respondents from higher performing schools cited more ways that the intervention plan impacted teaching than the lower performing school respondents. Mainly noted were faculty enthusiasm and a focus on struggling readers as a result of the intervention plan.

Comments Concerning Principal Leadership:

Further information clarifying the nature of principal leadership in ARI schools is provided through open-ended responses from teachers, reading specialists, and higher education partners. Responses are grouped into two categories:

- Building and classroom level support

- How the principal supported or impeded ARI implementation

Building and Classroom Level Support: In relation to the survey questions about building level support provided by the principal, teachers' responses can be sorted into the six categories and illustrative comments that follow. The most frequently cited categories were resources and materials and changes in scheduling. Other categories included support for workshops and professional development opportunities, general encouragement and leadership, classroom involvement and teacher supervision, little or no support. Kinds of support provided were not differentiated by high and lower performing schools.

The most noteworthy distinction between the higher and lower performing schools was found in the "provided little or no support." Teachers in lower performing schools were more likely to indicate that the principal provided little or no building or classroom level support than teachers in higher performing schools. Teachers in lower performing schools were less likely to say that the principal facilitated the acquisition of resources and materials and were more likely to indicate that the principals provided general encouragement and teacher support. It should be noted that the reports of "little or no building level support" in Cohort B are largely attributable to teachers' ratings from three schools.

When asked about the kinds of support that the principal provided to the school and classrooms, the reading specialists paralleled those of teachers. Monitoring program delivery and facilitating effective schedules were the most frequently cited ARI support from principals as identified by the majority of reading specialists in higher performing schools in both cohorts. Reading specialists in higher performing schools also indicated that the principals' efforts to facilitate professional development and their acquisition of resources and materials demonstrated their support of ARI implementation.

How the Principal Supported or Impeded Implementation of ARI: When asked how the principal supported or impeded implementation of ARI, the most frequently cited method by **teachers** was the provision of affirming support. Teachers' responses did not discriminate between higher or lower performing schools and can be categorized into the six areas that follow: provided materials or resources, provided professional development, provided affirming or moral support, provided supervision or administrative support, worked directly with students, and impeded progress.

When asked about ways in which the principal supported or impeded implementation of ARI, responses from **the reading specialists** fell into the same categories as those of the teachers. According to the reading specialist, the most frequently cited ways in which the principal supported ARI in higher performing schools was through affirming support and program leadership. Only one reading specialist from a lower performing school indicated these as mechanisms for support. Principals in both higher and lower performing Cohort B schools were identified as supporting implementation by providing program supervision (monitoring) and facilitating effective scheduling.

Principals in several higher performing schools were credited by the reading specialists with supporting the program by facilitating professional development. No principals in lower performing schools were so identified.

Analysis of responses from **higher education partners** concerning the leadership of the principal points to direct involvement of the principal as a difference between higher and lower performing schools. Comments relative to the higher performing schools noted that the principal supported ARI by observing and praising staff and students, providing for professional development and purchase of materials, requiring that every student have a reading book, checking student writing, listening to

students read, expecting to see modules implemented, and using faculty/grade level meetings to discuss ARI.

The above is in contrast to the higher education partner's remarks from a lower performing school, which pointed to a lack of leadership or faculty support and hindrance of important decisions. Responses were similar for higher and lower performing schools relative to daily schedule changes that included more reading time, help from the reading specialist for struggling readers, schedule changes such as two-hour language arts block and math block at a middle school and teaching reading early in the day.

Findings: Higher Education Partner

Introduction:

The higher education partnership is unique to the Alabama Reading Initiative since no other reading initiative throughout the United States contains this component. The higher education partnership brings reading educators from colleges and universities in Alabama into partnership with K-12 ARI schools to help affect research-based change in beginning reading, expansion of reading power, and effective intervention through the Alabama Reading Initiative model.

Twenty-one of Alabama's 32 colleges and universities were represented among the 64 higher education partners working with the 81 Literacy Demonstration Sites in Year 2. Information from the survey responses obtained from 43 higher education partners and 49 reading faculty who completed the surveys are detailed below relative to the impact of serving as a higher education partner (1) on the elementary, middle, or high school; (2) on the educator as a professional; and (3) on the college/university pre-service curriculum. Positive consequences and needs also are addressed. (It is noted that 36 of the reading faculty who completed surveys were higher education partners).

Higher Education Partner Participation in Summer Institutes:

Higher education partners reported their own participation in Summer Institutes. Of the 44 partners responding to the survey, from 18% to 41% either did not respond to elements of this item or indicated no participation. From 39% to 57% attended a module for the entire length of the presentation. Examination of full-time attendance differences among the higher and lower performing schools in each cohort revealed no apparent patterns associated with school outcomes.

Higher Education Partner Frequency of Visits:

Principals, reading specialists, and teachers indicated the frequency with which higher education partners visited their school. Respondents were asked to choose from among the selections shown in the table below. The majority of principals (60%) and reading specialists (58%) reported that the higher education partner visited their school at least once a month. The majority of teachers (63%) reported visits of less than once a month. No apparent pattern was revealed associating frequency of higher education partner visits with school outcomes.

Reported Frequency of Higher Education Partner Visits to School

Frequency of Visits	Principal Reports	Reading Specialist Reports	Teacher Reports
Every day	0	0	15
A few times per week	2	1	55
Once a week	4	3	56
A few times per month	10	15	193
Once a month	25	23	223
A few times per year	25	25	722
Once a year	2	3	105
Never	0	0	84
Total	68	72	1453
No Response	9	2	407

Higher Education Partner Activities and Helpfulness:

The following table indicates mean responses among the higher and lower performing groups for Cohort A and Cohort B. The number of reading specialist, principal, and higher education partner data points among the lower performing group in

Cohort A consisted of only one or two respondents. Differences were not calculated since they may not be meaningful.

Scale Means for Ratings of Higher Education Partner Frequency of Activities and Helpfulness

		Cohort A			Cohort B		
Survey	Rater	Higher Performers	Lower Performers	Difference	Higher Performers	Lower Performers	Difference
Higher Ed Partner Frequency of Activities	Principal	29.00	28.50		29.13	30.86	-1.73
Higher Ed Partner Frequency of Activities	Reading Specialist	28.67	18.00		25.14	25.14	0
Higher Ed Partner Frequency of Activities	Higher Ed Partner	27.00	*		24.50	26.00	-1.50
Higher Ed Partner Frequency of Activities	Teachers	20.61	18.94	1.67	21.52	22.15	-.63
Higher Ed Partner Helpfulness	Principal	29.00	18.50		27.73	28.71	-.98
Higher Ed Partner Helpfulness	Reading Specialist	28.33	18.00		24.64	21.71	2.93
Higher Ed Partner Helpfulness	Teachers	20.04	18.13	1.91	19.92	19.24	.68

*No responses were received in this group from higher education partners.

Principals, teachers, and higher education partners in Cohort B indicated a somewhat greater frequency of ARI activities in lower performing schools. However, reading specialists and teachers in Cohort B higher performing schools tended to rate higher education partner helpfulness more positively than did their counterparts in lower performing schools. No correlations greater than .20 were found between ratings of higher education partners and school outcomes. Results regarding ratings of higher education partners must be interpreted with caution due to a high incidence of missing data, especially on teacher surveys. This missing data to some extent is a function of what appears to be a lack of visibility of higher education partners among teachers. This conclusion is based on results reported in the previous section concerning higher education partner visits.

Comments Concerning The Higher Education Partner:

To better understand the role of the higher education partner in ARI, teachers, principals, and reading specialists responded to open-ended questions related to

- Activities of higher education partners
- Impact as a professional and service provider to the local school systems.

Activities of Higher Education Partners in Higher and Lower Performing Schools:

The 44 higher education partners who completed the survey reported that they performed a variety of activities in the ARI schools and commented that the partnership worked well. Activities described by the partners in higher performing and lower performing schools differed.

In **higher performing schools**, the higher education partners said that they conducted professional development sessions; modeled instruction; arranged for workshops; worked with struggling readers; worked one-on-one with teachers; trained new teachers in ARI module content; consulted with the principal, reading specialist and teachers; secured business support; and conferred with parents. In **lower performing schools** the higher education partners reported that they mostly did not directly interact with students and teachers in terms of instruction but served more as a consultant. The terms "sounding board, conversations about literacy, collaborative role" were used to describe aspects of the partnership that worked well.

When self-reporting about aspects of the higher education partnership that did not work well, one partner in a higher performing school cited distance from the university, and three others noted personnel issues such a lack of a reading specialist, overwhelmed teachers, and the relationship between the principal and reading specialist. All others felt that the partnership worked well.

- Response of Principals to Activities of Higher Education Partners

Eighteen of 21 principals in the higher performing schools identified specific aspects of higher education partnering that worked well at their schools. Professional development/workshops were cited most often followed by working with students, providing materials, working with teachers, and visiting classrooms.

Comments from principals in lower performing schools did not focus on specific activities but addressed the partner in general terms: compatible, eager to solve problems, not available, gave useful information and ideas, presented one workshop, knew if we needed anything partner would help. In one instance in a lower performing school, the partner visited on a bi-monthly basis, conducted workshops, coached teachers, and provided materials; but lack of leadership at the building level inhibited progress, according to the survey.

- Response of Teachers to Activities of Higher Education Partners

Teachers cited the most successful aspect of the higher education partnership as the partner's availability to help coach teachers, provide professional development, and model instruction in the classroom. A greater proportion of teachers in higher performing schools than in lower performing schools indicated the value of the partner as a model for instruction and as a resource for professional development. Teachers responses can be categorized into the following five areas: provision of professional development, modeling of instruction, partnership with the reading specialist, availability as a consultant for faculty, and providing resources.

The most frequently cited aspects that did not work well according to the teachers' surveys were the lack of presence and the lack of activity with faculty. The majority of comments indicated no problems with and appreciation for the higher education partner.

- Response of Reading Specialists to Activities of Higher Education Partners

Reading specialists' comments substantiate those of the higher education partner, the principal and the teachers. In the higher performing schools, they cited professional development/workshops most frequently as aspects that worked well, then conferring, and modeling. One reading specialist summed it up by saying, "The higher education partner conferring with teachers and the reading specialist was very helpful. She gave several workshops...offering ideas directly related to concerns at hand. This is a wonderful part of the ARI program." Another said, "Our partner was outstanding. She rolled up her sleeves and worked, as well as guided us. She was an absolute jewel! She pointed the way and we followed gladly."

In the lower performing schools, descriptions of what worked well are less specific and indicated limited involvement or were negative: "was willing to come and help" and "...only one professional workshop given by her" and "our first partner did not work out."

In terms of what did not work well, all but two respondents in the higher performing schools stated the higher education partnership worked well but expressed a desire to have the partner in the school more often. "Twice a month would be much better" was mentioned several times. One reading specialist cited the problem of distance between partner and schools as a problem and another stated that she didn't work with teachers much.

In the lower performing schools a lack of involvement and not coming enough, "not interested in hands-on experience," were noted. One specialist stated, "Help must be more than a long distance cheerleader. This is especially true if the principal is not an instructional leader."

Impact on the Higher Education Partner as a Professional and Service Provider to the Local School Systems: When asked about the impact of the higher education

partnership on them as professionals, no difference exists between comments of those who worked in higher performing schools and in lower performing schools. Comments were extremely positive in all instances except for one where the partner stated that the "experience was frustrating because of being a change agent in an unknown school culture."

According to the surveys, the higher education partnership has increased communication and cooperation between K-12 schools and institutions of higher education and enhanced professional growth at all levels. higher education partners who responded to the survey clearly cited numerous benefits of the partnership to the schools and to the colleges/universities in which they teach. They feel that they have grown as professionals and that their students have, in turn, grown in the pre-service arena. Professional benefits in terms of better preparation of pre-service teachers is cited frequently, "It has helped me better prepare new teachers." and "I can share the practical experiences that I had in a middle school on Monday with a content reading class on Wednesday." Another said, "Having a higher education partner provided teachers with professional instruction, new ideas, support, and leadership." and "The higher education partner is an important aspect of the program." One respondent referred to the higher education partnership as the strongest aspect of ARI. Another called the partnership relationship "good for all stakeholders."

Others commented that working with faculty, administration and struggling readers created an awareness of the benefits to schools and to the partner as a higher education faculty members. These opportunities gave them what they termed "a connection to the real world" and "a reality check." The partnership helped them "continue to research and to grow in the area of literacy," and provided insights into how school faculties cooperate to implement programs, opportunities to asses students' reading abilities, a chance to model and use strategies with K-12 students, additional

knowledge about methodology that works, and opportunities to work with teachers and students in meaningful ways and network with reading professionals throughout Alabama. Specific comments follow.

Additional Comments: Teacher, principal, reading specialist and higher education partner self-reports point out the importance of the higher education partner to the success of the ARI when the higher education partner provides professional development, models strategies, works with students, and maintains direct contact with faculty and staff. Differences exist in higher and lower performing schools in the helpfulness of the higher education partner, especially as reported by the reading specialist. Higher education partners also have described numerous benefits of ARI in terms of college/university course content and site-based experiences for pre-service teachers.

Teacher survey results point to the importance of ongoing professional development (workshops, modeling of instruction, and other activities) related to ARI training – a role for which the higher education partner was praised. This innovative partnership has been reported to be successful in many ways.

Higher Education Institutional Reward System: One partner said that ARI has asked a lot of services from the higher education partner but little is done to reward him/her. Another said, "It detracts from my ability to plan for my classes and do research that is required." Other comments focused on the need for ongoing professional development in relation to ARI in the schools.

The time factor is substantiated in the principals' surveys. In response to what did not work well in relation to the higher education partnering, time that the partner was available to work in the school was identified as the biggest negative consequence by principals. Nine of 14 principals from higher performing schools said that time was a problem. They felt that the effectiveness of the partnership would have been increased

if the partner had more time available to work in the school. Two principals from higher performing schools indicated that there were no problems; one said that nothing worked well, one cited distance of the partner from the school, another would have liked to see better coordination in specialization areas.

Of the seven lower performing school principals who responded, four indicated that they would have liked to see the higher education partner more often. One principal said that the partner was not familiar with her role; one indicated that the partnering did not work, and another felt that more coaching and more positive feed back would have been beneficial. In response to other survey questions, several principals again focused on the need for additional time for the partner to work in the school. Reading specialist and teacher comments indicate a desire to have the partner in the school more than once and month so that the partner can be even more supportive.

QUESTION 4: WHY ARE SOME ARI SCHOOLS MAKING MORE PROGRESS THAN OTHERS?

Summary of Findings for Question 4

The key elements that appeared to differentiate higher and lower performing schools were the principal's leadership in implementing ARI, the reading specialist's helpfulness, and the helpfulness of the higher education partner.

Detailed Findings for Question 4

The answer to why some schools are making more progress than others is suggested in the results obtained from the quantitative and the qualitative data. The program is a clear delineator between ARI and non-ARI schools. Even though some ARI schools experienced declining performance, the ARI schools as a group out-performed the state's non-ARI schools.

The Principal as ARI Champion

Other than the ARI training itself, data point to the principal as having the greatest impact on student outcomes. The principal's support of ARI correlated highly with outcomes. Examination of correlations between principal leadership scores and *Stanford 9* effect sizes revealed correlations of .20 or greater between leadership and one or more of the *Stanford 9* reading scores, regardless of the source of the ratings. These findings highlight the importance of principal leadership in promoting academic gains in ARI schools. In higher performing schools the principal facilitated the acquisition of resources and materials, supported professional development, and offered frequent encouragement. In lower performing schools little or no support was noted.

The Hands-On Reading Specialist

The reading specialist also made a difference. Reading specialist helpfulness correlations were found in excess of .20 in relation to Reading Comprehension and Total Reading on the *Stanford 9*. Furthermore, teacher and principal ratings of the helpfulness

of the reading specialist were higher in higher performing schools than in lower performing schools. Also, in lower performing schools, the reading specialist was more often regarded as one who held more of an administrative/supervisory role than in higher performing schools.

The large majority of teachers in Cohorts A and B considered the role of the reading specialist as important to the successful implementation of the ARI. Teachers reported that the reading specialist supported implementation of ARI and impacted reading abilities of struggling readers primarily through working directly with students as well as through their efforts to encourage and assist teachers. Teachers also recognized the role of the specialist in providing training, providing materials and resources, and providing program oversight or facilitating scheduling to enable optimal delivery of the ARI.

The Deeply Involved Higher Education Partner

Differences exist in higher and lower performing schools in the perceived helpfulness of the higher education partner, particularly as evaluated by the reading specialist. In high-performing schools, the higher education partner is perceived by the reading specialist as being more helpful than in lower performing schools. The higher education partners in higher performing schools tended to report more direct involvement with students, teachers, and reading specialists. Higher education partners in lower performing schools reported serving more as consultants.

Key Process Elements

Process variables that differentiated higher and lower performing schools, as defined earlier by teachers, include:

- more hours of professional development during the school year,
- an increase in student reading time,

- an increase in reading activities,
- greater attention to reading strategies,
- more assessment of reading ability,
- increased awareness of the needs of readers,
- greater confidence and enthusiasm regarding reading instruction, and
- increased focus on struggling readers.

PART 4: PRE-SERVICE COMPONENT

Key Issues

The long-range solution to Alabama's pursuit of 100% literacy lies, at least in part, with teacher preparation programs. Teacher preparedness to instruct all public school students to learn to read, to expand their reading power, and to provide intervention for struggling readers depends in part on the preparation of teachers in pre-service programs.

Evaluation Methods

Surveys were distributed to both higher education partners and higher education reading faculty at 32 public and private institutions of higher education in Alabama. These surveys elicited responses to open-ended items related to pre-service curricula and practices or experiences that impact the delivery of pre-service programs. Surveys returned included 44 higher education partner surveys and 49 higher education faculty surveys. Responses were analyzed in order to answer the evaluation questions that follow.

QUESTION 1:	**TO WHAT EXTENT ARE THE ELEMENTS OF ARI REFLECTED IN PRE-SERVICE TEACHER EDUCATION PROGRAMS THROUGHOUT ALABAMA?**
QUESTION 2:	**WHAT ARE FACTORS ARE RELATED TO CHANGE IN PRE-SERVICE TEACHER EDUCATION PROGRAMS?**

Summary of Findings for Questions 1 and 2

- All higher education partners and reading faculty indicated changes in course content that reflected content of ARI learning modules and published documents.
- All higher education partners and reading faculty who responded to surveys indicated that current course syllabi incorporated the new reading standards adopted by the Alabama State Board of Education in December 1999. Those standards reference specifically *Knowledge and Skills Teachers Need to Deliver Effective Reading Instruction*, a document developed by the ARI and published by the Alabama State Department of Education in February of 1998.
- Reading faculty noted the positive benefits to student teachers and field experience students who serve in ARI schools where they can observe and participate in research-based, effective instruction.
- Representatives from higher education faculty served in various capacities with ARI: Reading Panel, higher education partnership, module development and revision, summer institute presentations, advanced study committee.

Detailed Findings for Questions 1 and 2

Impact on College/University Pre-service Curriculum

Higher education partners and faculty most frequently cited inclusion in course content of the ARI modules and ARI published documents, *Report on the Review of Research* and *Knowledge and Skills that Teachers Must Know and Use to be Effective Teachers of Reading*. Eleven responses identified a range of ways in which information and research from ARI was incorporated in coursework: "Much of the content that was presented at Summer Institutes has been incorporated into various courses" and "we are...aligning course content sequentially so our pre-service teachers have systematic instruction and application of the knowledge and skills...that are supported by ARI research."

All higher education partners and reading faculty who responded to the survey indicated that current reading course syllabi at their institution incorporated the new reading standards adopted by the Alabama State Board of Education in December 1999. Almost all of the respondents said that major changes were made in course content; only three noted minor changes; and one, no change since course content paralleled ARI. These standards reference the *Knowledge and Skills that Teachers Must Know and Use to be Effective Teachers of Reading*, developed by ARI and published by the Alabama State Department of Education.

Second most frequently mentioned change was the effectiveness of pre-service teaching experiences in Literacy Demonstration Sites where students could observe, complete field experiences, and internships, and become a part of research-based effective practice. They also noted the effectiveness of using examples from schools to illustrate points in the college/university classroom, having a common vocabulary, and "promoting the goals of ARI where ever graduates end up teaching" as having a positive impact.

As one higher education partner said, visiting classrooms and tutoring struggling readers "allowed me to provide (my students) with methods/strategies that worked well with specific grade levels and content areas." Another noted, "They see us as partners - not adversaries." Additional comments included the following:

- More hands on experiences with informal assessment
- Balance between decoding and holistic components
- More focus on research base
- Strengthening of lecture components with ARI materials
- More aware of teachers' needs and concerns
- Validates our reading program
- Best practices demonstrated by ARI trained teachers to pre-service students
- Relationship between theory and practice seen by students
- Gives students a more specific way to teach reading in the classroom
- Provides a professional resource
- Helps to unite and give a common articulation
- Professionalizes the field – teachers teaching teachers
- Makes all more aware of what the research says
- Consistency in terminology, philosophy, and methods
- Provides more effective teachers as cooperating teachers for our student teachers
- Teacher educators and teachers becoming part of a strong professional network
- Focuses attention on the centrality of literacy to all programs
- Serves as a foundation for change in reading instruction
- Made programs from various universities more consistent and improved some programs and made them more accountable

- Developed partnerships between universities and schools to provide more hands on experiences in reading instruction
- Improved interest in graduate education because of partnership

Two respondents said that they already emphasized what ARI does.

Impact on Faculty Knowledge Base

Of the 49 survey respondents, 44 of the higher education faculty who responded to the survey indicated that they were very familiar with ARI, and they demonstrated knowledge of ARI and its goals: 100% literacy, professional development that is research-based, intervention for struggling readers, research-based reading instruction, expand reading ability of those already at grade level or above.

Nineteen higher education partners and faculty indicated involvement in presenting some or all of the seven learning modules at the Summer Institutes, with some professors presenting more than one learning module. Ten faculty members indicated that they have presented ARI modules to K-12 schools during the school year from 1998-2000. Thirteen reported being involved in developing and/or revising one or more of the seven modules, and three have served a module development and/or revision chairpersons.

Limitations and Obstacles to Higher Education Participation in Sustained Activities to Local School Systems

- Faculty Workloads:

Workload concerns and competing time demands were cited by higher education partners and higher education reading faculty as negative consequences to serving ARI. "The workload is unbelievable" and "These efforts are over and above our 'normal' workload" were voiced by many. One respondent referred to the impact of time by citing specific involvement, "time spent over and above the university work load to travel to Montgomery for training, serve more than one school effectively, and serve on additional

ARI committees." That "universities need to recognize the importance of the partnership in accelerating reading ability and provide load time to work with the schools" also was reflective of most comments.

- Retention:

One respondent voiced major concerns because of difficulty hiring and retaining reading faculty since some reading faculty have chosen to become reading specialists.

PART 5: LIMITATIONS OF THE EVALUATION

Although the evaluation addresses many aspects of change that took place at the 81 LDSs, several limitations exist to the design. These limitations, outlined below, should be kept in mind when interpreting the results of the study.

1. The survey instruments were designed for self-reporting of principals, teachers, reading specialists, and higher education partners. Factors such as the levels of implementation of ARI within schools were not validated by independent observations.
2. This evaluation addresses only those components of student, teacher, reading specialist, principal, higher education partner, and pre-service outcomes specified in the evaluation plan and reported on surveys or conveyed through test data. Any or all of the LDSs may have additional results that could in some cases reveal a more complete picture.
3. Understanding the impact of ARI on student performance is further complicated by the substantial variability in the performances of schools prior to the implementation of ARI. That is, in some schools student performance was substantially higher than in others. This may allow for some schools to make more substantial gains on standardized measures of achievement due to influences such as statistical regression to the mean.
4. Evaluators were unable to determine an appropriate analytical strategy for incorporating the K, 1 and 2 statewide reading assessments as part of the evaluation of ARI. The tests currently used to assess early literacy skills of students in kindergarten through second grade are not designed to assess ongoing development of reading ability. Therefore, the evaluation does not present information concerning the progress of students in the first three

years of school, when progress might be expected to occur at a greater rate than in upper grades. Recommendations for using the existing assessment or alternative procedures for gathering literacy outcome data in kindergarten through third grade are included in the Recommendations section of this report.

5. Cross-sectional and longitudinal effects of ARI documented using the norm-referenced *Stanford 9* may be suppressed due to a statistical artifact associated with using norm-referenced tests to compare groups. Items on norm-referenced tests are selected to maximally discriminate between individuals (thereby yielding the normal distribution). Because such tests tend to maximize the spread of scores within groups of students (such as students within Literacy Demonstration Sites), it is more difficult to detect differences between groups using such tests (Hashway, 1976). Effect size estimates are particularly influenced by the maximization of the spread of scores. For any given difference in mean NCE scores between groups, as the amount of spread within groups increases the effect size will decrease.
6. The higher and lower performing schools were selected using a set of criteria chosen by the evaluators. Different sets of criteria might result in the selection of other groups of schools. Therefore, conclusions regarding differences between higher-performing and lower-performing schools should not be generalized beyond the school classification model used for this evaluation.

Importance of Interpreting *Stanford 9* Evaluation Results in Context

This evaluation of the Alabama Reading Initiative must be interpreted in light of the context of educational accountability in Alabama. The Alabama Educational Accountability Act mandates that schools in Alabama will have the majority of students

performing at or above grade level or face potential sanctions. The *Stanford 9* is the mandated instrument for measuring educational accountability in third through eleventh grade across all public schools. The State has defined the fifth stanine as the criterion level of performance at or above grade level. Therefore, schools that do not have more than 50% of their students falling at or above the fifth stanine are required to make prescribed levels of progress or face sanctions. As a result of this law, teachers and administrators in all Alabama elementary and secondary schools are under considerable pressure to facilitate student performance on the *Stanford 9*, regardless of the school's involvement in ARI.

Schools that are not performing up to expectations in terms of the accountability criteria are provided support by the Alabama State Department of Education. Three problems emerge as a result of this state and local emphasis on raising *Stanford 9* performance scores. First, efforts to compare schools that participate in the ARI with those that do not leads to comparisons between "competing conditions" rather than between "treatment" and "control" schools. This reduces the likelihood of detecting substantial differences between the ARI and non-ARI schools. Second, as the Educational Accountability Act has been in effect since 1996, schools that do not have more than 50% of students at or above the 5th stanine have been implementing strategies and curriculum reforms in order to comply with accountability criteria. Therefore, monitoring cross-sectional changes on the *Stanford 9* is confounded by fluctuations and gains that are attributed to a history of school-based reform efforts. Third, the curriculum reforms that are being implemented in some schools in order to enhance *Stanford 9* performance in weak areas such as mathematics and language may detract from attention to reading and writing processes and outcomes of the ARI. For example, if an ARI school is on Academic Alert due to student performance deficits in mathematics, teachers and administrators in that school may be less inclined to focus on

literacy objectives if doing so might reduce instructional time devoted to improving mathematics performance.

PART 6: RECOMMENDATIONS

Recommendations Regarding Delivery of the Alabama Reading Initiative

1. The principal's leadership role in supporting and facilitating the implementation of the ARI in Literacy Demonstration Sites was related to gains in *Stanford 9* reading scores. Principals should therefore receive sufficient support, direction, and encouragement regarding their leadership role with the ARI. Teachers reported enthusiasm for principals who provided conditions which enabled successful implementation and who were visible and engaged with teachers and students in implementing ARI. Teacher attitudes toward ARI appear to be, in part, shaped by the attitudes of the principal. If a change in principal occurs, consideration should be given to the candidate's attitudes toward the ARI when making selection decisions. It is important that principals be well-versed in the content, methodologies, principles, and objectives of ARI. It may benefit the state to establish a listserv or other communication system for principals in Literacy Demonstration Sites to share ideas and recommendations for supporting the ARI within their schools. Different needs appear to exist in elementary, middle/junior high, and secondary schools. Therefore, separate technology-based discussion groups may be needed for each type of school.
2. The visibility and involvement of the reading specialist and higher education partner influenced teachers' attitudes toward the helpfulness of these individuals. Therefore, to the extent possible, efforts should be made to: (a) clearly understand the roles and responsibilities of these individuals and communicate this information to all school personnel and ARI support staff, (b) increase the visibility of these positions within Literacy Demonstration Sites, (c) support the direct involvement of these persons with classroom teachers, and (d) increase

the number of reading specialists available to directly work with teachers and students.

3. Higher education faculty who participate as partners or as reading support personnel are often confronted with difficulties associated with inadequate or competing reward structures for their involvement in ARI. That is, for higher education partners who valued their work in the Literacy Demonstration Sites, there were potential unanticipated costs associated with such involvement. For example, the partners who spent a substantial amount of time in schools indicated that they might be compromising their opportunities for reward within their higher education institutions because they were not able to spend as much time on scholarship as were their colleagues. To the extent that the higher education partnership is to be preserved as an aspect of ARI, higher education faculty must be provided with adequate release time and assurances needed to enable a sufficient commitment of time and energies to the assigned Literacy Demonstration Site. Furthermore, strategies are needed to increase the direct involvement of higher education partners in assisting classroom teachers.
4. Teachers in Cohort A indicated that continued professional development during implementation would help sustain their enthusiasm for improving their reading instruction. Therefore, we recommend that systematic professional development designed to meet the different needs of second and third year Literacy Demonstration Sites be included as an aspect of the ARI. The ability to sustain enthusiasm for ARI appears to be influenced by the encouragement and attention that teachers receive from stakeholders. Therefore, we recommend that efforts be made to sustain local and state recognition of efforts made by Literacy Demonstration Sites to raise literacy levels within schools beyond the first year of implementation.

Recommendations Concerning Evaluation of the Alabama Reading Initiative

1. As a norm-referenced survey battery achievement test, the *Stanford 9* is not the most appropriate measure for monitoring progress toward 100% literacy in Literacy Demonstration Sites in Alabama. The evaluation team understands that the nature of the statewide accountability model currently necessitates the use of this test as a primary index of gains. Nevertheless, there is a need to identify alternative measures that more directly assess changes in literacy. Informal Reading Inventories or other diagnostic reading assessments provide criterion-referenced outcomes that may be more appropriate for documenting the nature and level of change in literacy. This concern is also relevant in the discussion of outcome criteria such as "student reading at or above grade level". Norm-referenced tests, such as the *Stanford 9*, are not constructed to permit clear delineations of such performance levels. The criteria used in this evaluation were adopted, to a large extent, from conventions used as part of the Alabama accountability system. Therefore, if an alternate outcome measure is to be considered for the ARI evaluation, such an instrument must be able to provide criterion-referenced judgments (such as above or below grade level).
2. The tests currently used to assess early literacy skills of students in kindergarten through second grades are inappropriate for use as tools for evaluating gains associated with ARI. Therefore, the evaluation report does not present information concerning an important goal of the Alabama Reading Initiative. Most subtests of the current literacy assessment have sufficient reliability in order to be used to determine changes in scores during a given academic year (i.e., September through April). We support the use of criterion-referenced measures that are designed to document progress of students toward reading at or above grade level. Implementing fall and spring testing using the individual or group

form of the criterion-referenced reading assessment administered in kindergarten through grade two and/or linking these measures psychometrically would permit the use of available instruments for this purpose.

3. The current evaluation documented changes in groups of students followed across specific grade levels. This was done in order to ascertain whether approximately the same group of students within a school made improvement. However, this evaluation did not limit this longitudinal comparison to exactly the same group of students across time. Therefore, if a school experienced substantial changes in the student body from one grade level to the next, longitudinal comparisons for that school will be confounded with cross-sectional changes. A more desirable approach is to conduct a true longitudinal evaluation in which the same students are tracked across time. Changes across grades for such matched data would provide the most direct measure of the impact of ARI on student achievement across time. Therefore, we recommend longitudinal analysis of *Stanford 9* data be conducted as an integral part of the Year Three evaluation of the Alabama Reading Initiative.

4. A large percentage of principals, teachers, reading specialists, and higher education partners reported that the Alabama Reading Initiative had substantial positive impacts on student literacy. Findings based on the *Stanford 9*, while tending to support the efficacy of the Alabama Reading Initiative, were somewhat less positive and consistent. The evaluators believe that increases in reading time, reading interest, and reading skills did occur in ARI classrooms, including some classrooms in schools that did not demonstrate gains on the *Stanford 9* reading subtests. We recommend that alternate measures be considered that: (a) are more sensitive to desired changes in literacy, (b) are administered on a

more frequent basis, and (c) provide more diagnostic information that can be used by the teachers, schools and state for improving reading instruction.

5. Aggregating data across 81 schools, more than 1,000 teachers, and more than 20,000 students will invariably result in the loss of important information that accounts for differences in outcomes between schools and between teachers within schools. While this evaluation attempted to account for differences in performance of groups of schools, an alternate methodology appears warranted. We recommend that case studies of particularly effective schools be conducted in order to identify common conditions that support efficacy. Based on the current study, specific issues that should be examined as part of such case studies include principal leadership, the role of the reading specialist, the nature of resources available to teachers, the role and involvement of the higher education partner, and the level of support and encouragement provided to teachers implementing ARI. We recommend that such case studies sample a range of school types and demographic contexts in order to gain a representative picture.
6. Non-academic outcomes such as disciplinary referrals, special education referrals, and library circulation rates provide important collateral evidence of the efficacy of the Reading Initiative. Such collateral outcomes should be included as a substantial feature of subsequent evaluations.